

**JOÃO MARQUES SALOMÃO**

**IMPLEMENTAÇÃO DE MÉTODOS BIOMÉTRICOS BI-MODAIS  
BASEADOS EM FUSÃO DE CARACTERÍSTICAS PARA  
RECONHECIMENTO DE INDIVÍDUOS**

Tese apresentada ao Programa de Pós-Graduação em Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do Grau de Doutor em Engenharia Elétrica, na área de concentração em Automação.

Orientador: Prof. Dr. Ailson Rosetti de Almeida.

Co-Orientador: Prof. Dr. Evandro Ottoni Teatini Salles.

VITÓRIA  
2007

Dados Internacionais de Catalogação-na-publicação (CIP)  
(Biblioteca Central da Universidade Federal do Espírito Santo, ES, Brasil)

---

S173i Salomão, João Marques, 1956-  
Implementação de métodos biométricos bi-modais baseados em fusão  
de características para reconhecimento de indivíduos / João Marques  
Salomão. – 2007.  
133 f. : il.

Orientador: Ailson Rosetti de Almeida.

Co-Orientador: Evandro Ottoni Teatini Salles.

Tese (doutorado) – Universidade Federal do Espírito Santo, Centro  
Tecnológico.

1. Sistemas de reconhecimento de padrões. 2. Inteligência artificial. 3.  
Redes neurais (Computação). 4. Visão por computador. 5. Processamento  
de imagens. I. Almeida, Ailson Rosetti de. II. Salles, Evandro Ottoni  
Teatini. III. Universidade Federal do Espírito Santo. Centro Tecnológico.  
IV. Título.

CDU: 621.3

---

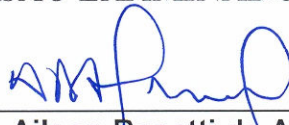
**JOÃO MARQUES SALOMÃO**

**IMPLEMENTAÇÃO DE MÉTODOS BIOMÉTRICOS BI-MODAIS  
BASEADOS EM FUSÃO DE CARACTERÍSTICAS PARA  
RECONHECIMENTO DE INDIVÍDUOS**

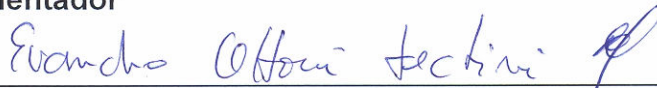
Tese submetida ao programa de Pós-Graduação em Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisição parcial para a obtenção do Grau de Doutor em Engenharia Elétrica na área de concentração em Automação.

Aprovada em 24 de agosto de 2007.

**COMISSÃO EXAMINADORA**



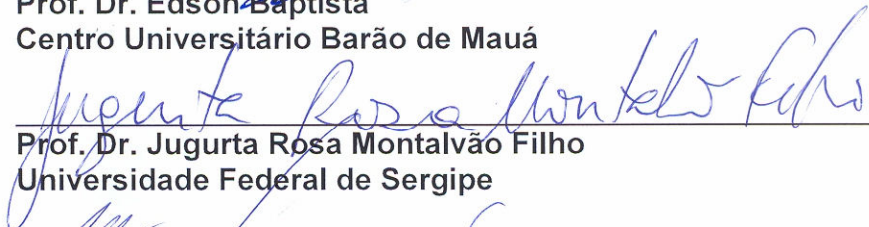
Prof. Dr. Ailson Rosetti de Almeida  
Universidade Federal do Espírito Santo  
Orientador



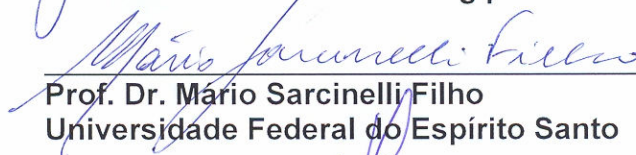
Prof. Dr. Evandro Ottoni Teatini Salles  
Universidade Federal do Espírito Santo  
Co-Orientador



Prof. Dr. Edson Baptista  
Centro Universitário Barão de Mauá



Prof. Dr. Jugurta Rosa Montalvão Filho  
Universidade Federal de Sergipe



Prof. Dr. Mário Sarcinelli Filho  
Universidade Federal do Espírito Santo



Prof. Dr. Teodiano Freire Bastos Filho  
Universidade Federal do Espírito Santo

# Dedicatória

À minha esposa, Rachel, aos meus filhos Juliana, Mariana e Martinho pelo incentivo, carinho, compreensão e paciência que tiveram comigo ao longo dos vários anos que me dediquei a esta Tese.

A meus pais, Olma (in memorian) e Martinho, que me deram a vida e que, apesar da falta de escolas nas áreas rurais e das dificuldades apresentadas naquela época, fizeram todos os esforços possíveis para viabilizarem meus estudos até a adolescência e continuaram me incentivando a estudar na minha fase adulta.

Às minhas tias Olga (in memorian), Janete (in memorian), Vivi e aos meus primos que, apesar da numerosa família e das dificuldades financeiras, carinhosamente me acolheram e me apoiaram quando aqui cheguei vindo do interior com uma visão curta e limitada do mundo.

# Agradecimentos

Inicialmente, gostaria de agradecer a meus orientadores, professores Dr. Ailson Rosetti de Almeida e Dr. Evandro Ottoni Teatini Salles, sem os quais esta tese não seria possível. Suas orientações foram fundamentais para que o trabalho chegasse a bom termo. Seus entusiasmos e suas sugestões me deram o ânimo necessário para continuar trabalhando, mesmo nos momentos mais difíceis. Ao professor Dr. Jugurta Rosa Montalvão Filho que, com sua visão privilegiada, me fez ver neste trabalho mais contribuições do que eu imaginava. Ao professor Dr. Edson Baptista, pelas críticas e sugestões que fez a este trabalho para torná-lo ainda mais significativo. Ao professor Dr. Mário Sarcinelli Filho, pelas sugestões e pela minuciosa leitura que fez deste manuscrito, e cujas correções permitiram melhorar muito a apresentação e redação final desta tese. Ao professor Dr. Teodiano Freire Bastos Filho, pelo incentivo, pelas críticas e sugestões oferecidas ao longo deste trabalho. Ao colega professor Dr. Hans Rolf Kulitz, pelo incentivo e pelo valioso conhecimento que me proporcionou durante mais de 20 anos de convívio, o que me permitiu chegar com segurança ao meu objetivo, e aos colegas professores M.Sc. João Checon Neto, M.Sc. Luis Eduardo Martins de Lima e M.Sc. Nilson Santos Marcelus, pelo apoio e incentivo durante toda a minha caminhada. Aos colegas M.Sc. Rober Marconi e M.Sc. Klaus Fabian Coco, com os quais compartilhei muitos momentos de estudo e de pesquisa ao longo do desenvolvimento deste trabalho. Finalmente, um agradecimento especial a todos os colegas professores da Coordenadoria de Eletrotécnica do CEFET-ES por proporcionarem a liberação das minhas atividades para que eu pudesse dedicar-me a esta pesquisa.

**“Alguns homens vêem as coisas como são, e dizem ‘Por quê’?. Eu sonho com as coisas que nunca foram e digo ‘Por que não?’”.**

**George Bernard Shaw**

# Resumo

Esta pesquisa propõe uma forma de fusão de informações no nível das características para um sistema bi-modal de reconhecimento pessoal. As características biométricas e comportamentais consideradas na validação deste trabalho são obtidas das seqüências de vídeo da forma humana de caminhar e de imagens faciais. Ela sugere arquiteturas e algoritmos que permitem a implementação de reconhecimento e autenticação de indivíduos através da fusão de características faciais e da forma humana de caminhar, objeto de poucos trabalhos disponíveis na literatura. A escolha da abordagem através da fusão das duas características justifica-se pela suposição de que ela deve oferecer um melhor desempenho de classificação, robustez e segurança ao permitir a identificação noturna e à distância. Além disso, é considerada menos intrusiva do que todos os outros sistemas biométricos, por apresentar pouca ou nenhuma necessidade de colaboração da pessoa a ser identificada. A pesquisa, pela atualidade do tema e pelas poucas referências bibliográficas, inicia-se avaliando o desempenho do sistema de reconhecimento pessoal pela forma de caminhar ao se aplicarem, sobre as seqüências de vídeo, técnicas de extração e seleção de características de silhuetas baseadas em análise de componentes principais (*PCA - Principal Component Analysis*), análise de componentes independentes (*ICA - Independent Component Analysis*), transformadas Wavelets e proporção de variâncias (*PoV - Proportion of Variances*), em combinação com os classificadores baseados em distâncias Euclidianas, máquinas de vetores suporte (*SVM - Support Vector Machines*) e redes neurais com funções de bases radiais (*RBF - Radial Basis Function*). Utilizando-se bases de dados de domínio público, foi feito um estudo especial sobre a forma de caminhar objetivando uma avaliação inicial do seu desempenho em sistemas de reconhecimento pessoais, bem como a escolha de métodos e técnicas mais apropriadas para efetuar a fusão desta com as imagens faciais do indivíduo. Em seguida, obtiveram-se bases de dados próprias baseando-se em um cenário proposto que permitiu extrair as silhuetas de seqüências de vídeo da forma de caminhar e das imagens faciais correspondentes a cada seqüência. Avaliou-se o desempenho do sistema de reconhecimento quando se estabelece a fusão de ambas as características biométricas. Em uma etapa posterior, após a definição da arquitetura mais adequada, é proposta e implementada a fusão dos dois sistemas biométricos face-forma de caminhar sobre bases de dados de domínio público e proprietárias. Nesta fase final do trabalho, as implementações sugeridas utilizam-se das técnicas de extração e seleção de características baseadas na energia das silhuetas do caminhar, na proporção de variâncias (*PoV*) e em algoritmos de classificação baseados em redes neurais com funções de bases radiais (*RBF*). Os resultados apresentados permitiram avaliar as taxas de acertos, de falsa aceitação e falsa rejeição e o desempenho do sistema, quando considerados os vetores de características individuais das imagens faciais e de seqüências de vídeo do caminhar e da fusão das características de ambas, confirmando a validade e viabilidade de aplicação deste trabalho de fusão no nível das características no sistema bi-modal de reconhecimento pessoal.

Palavras-chave: Sistema bi-modal de reconhecimento pessoal; Fusão no nível de características; Seqüências de vídeo da forma de caminhar; Imagens faciais; Redes neurais com funções de bases radiais.

# Abstract

This research proposes a bi-modal personal recognition system based on the information fusion in the level of the characteristics (also called features). The biometric and behavioral features considered in the validation of our proposal are obtained from the video sequences of the human gait and face images. It evaluates some proposals of architectures and algorithms in the implementation of multiple biometric methods for recognition and authentication of individuals through the fusion of face features and the human gait, whose research subject is scarcely available in the scientific literature. The choice of the approach for the fusion of the two particularly considered information sources (not to be confused with multiple attributes of a feature vector from one information source) is justified by the supposition that it must offer a better classification performance, robustness and safety as it allows the night identification at a certain distance. Furthermore, it is considered less intrusive than all the other biometric systems since it presents little or no need for the collaboration of a person to be identified. The research, being an state of the art theme, has few bibliographical references and starts by evaluating the performance of the human recognition system through gait as it applies, on the video sequences, techniques of extraction and selection of silhouette features based on PCA (Principal Component Analysis), ICA (Independent Component Analysis), Wavelet transforms and PoV (Proportion of Variances), combined with the classifiers based on Euclidian distances, SVM (Support Vector Machines) and neural networks with RBF (Radial Base Functions). Using databases of public domain, a special study was carried out on gait aiming at an initial evaluation of its performance in human recognition systems, as well as the choice of methods and techniques more appropriate to make the fusion of this gait with the face images of the individual. As an improvement, on a second step, own databases were obtained based on an experimental setup that allowed the extraction of silhouettes from gait video sequences as well as face images corresponding to each sequence. This research also evaluated the performance of the recognition system when determining the fusion of both biometric features. Then, after the definition of the most adequate architecture, the fusion of two biometric systems face-gait on public domain databases, as well as our particular one, is proposed and implemented. In this final phase of the work, the proposed implementations use extraction techniques and features selection based on gait silhouettes energy, in the proportion of variances (PoV) and in the classification algorithms based on neural networks with radial base functions (RBF). The obtained results allowed the evaluation of the non-error rates, the false acceptance (FAR) and rejection rates (FRR) and the system performance, when the vectors of individual features of the face images alone, the gait video sequences alone, as well as the fusion of both are considered, confirming the validity and viability of the application of our proposal of fusion in the feature level in the bi-modal human recognition system.

**Keywords:** Bi-modal human recognition system; Fusion in the feature level; Gait video sequences; Face images; Neural networks with radial base functions.



# Lista de tabelas

Tabela 1 - Comparação das várias características biométricas (JAIN; ROSS; PRABHAKAR, 2004). Alto, Médio e Baixo são denotados por A, M e B, respectivamente. ....	39
Tabela 2 - Resultados dos classificadores sobre o total de vetores .....	74
Tabela 3 - Resultados por classe do classificador neural.....	75
Tabela 4 - Matriz de confusão: classificador K-NN .....	78
Tabela 5 - Matriz de confusão: classificador SVM.....	79
Tabela 6 - Desempenho SVM por sequência .....	80
Tabela 7 - Limiar percentual do vetor $r_m$ e redução do número de amostras.....	83
Tabela 8 - Matriz de confusão para o classificador SVM.....	86
Tabela 9 – Taxas de acertos: faces ( $Fa$ ), caminhar ( $Ca$ ) e a sua fusão ( $Fu$ ) em função da taxa PoV ( $r_m$ ), da dimensão do vetor de características ( $Cf$ ) e do tempo de processamento ( $Tp$ ).....	106

# Lista de figuras

Figura 1 - Arquitetura de um sistema de identificação biométrico.....	31
Figura 2 – Curvas representando a frequência de erros e acertos em função dos escores.....	36
Figura 3 – Curvas representando FAR, FRR e EER .....	37
Figura 4 - Exemplos de características biométricas (JAIN; ROSS; PRABHAKAR, 2004) .....	38
Figura 5 – Sistema clássico de reconhecimento facial .....	41
Figura 6 – Obtenção dos vetores de características de imagens faciais .....	42
Figura 7 – Vista de cima do cenário para a captura das seqüências de vídeo do caminhar.....	45
Figura 8 - Seqüência de silhuetas representando um “ciclo do caminhar” .....	45
Figura 9 – Extração de características com uso de abordagens baseadas em modelo .....	46
Figura 10 – Extração de características de silhuetas binárias.....	47
Figura 11 - Modelo ICA simplificado .....	54
Figura 12 - Decomposição Wavelet dos vetores de características .....	56
Figura 13 - Hiperplanos de separação para duas classes: a) separáveis b) não- separáveis .....	60
Figura 14 - Modelo de um neurônio artificial.....	63
Figura 15 - Estrutura de uma rede RBF.....	65
Figura 16 – Obtenção da imagem binária com a janela dinâmica de contorno .....	69
Figura 17 – Janela contendo as silhuetas interna e externa.....	70
Figura 18 - A abordagem utilizada.....	72
Figura 19 – Exemplos de quadros de seqüências utilizadas das bases de dados ...	73
Figura 20 - A arquitetura proposta .....	76

Figura 21 - Processo de treinamento em configuração de “validação cruzada” .....	77
Figura 22 – Síntese em diagramas de blocos da proposta utilizada .....	81
Figura 23 - Desempenho KNN x SVM para os 18 coeficientes independentes.....	85
Figura 24 - Arquitetura proposta.....	87
Figura 25 – Exemplos de quadros de seqüências da base de dados utilizada .....	88
Figura 26 – Exemplos de silhuetas de um ciclo e imagem ESC.....	89
Figura 27 - Resultados obtidos.....	90
Figura 28 – Estrutura de um sistema bi-modal: módulos de fusão face-forma de caminhar.....	96
Figura 29 – Exemplos de quadros das seqüências de vídeo e imagens faciais utilizadas .....	98
Figura 30 - Obtenção dos vetores de características da face e da silhueta .....	100
Figura 31 – Arquitetura com o esquema de fusão no nível de decisão .....	101
Figura 32 - Resultados obtidos: fusão no nível de decisão .....	103
Figura 33 – Arquitetura com o esquema de fusão no nível de características .....	104
Figura 34 - Resultados obtidos: fusão no nível de características.....	107
Figura 35 – Cenário proposto para a fusão face-forma de caminhar .....	110
Figura 36 – Exemplares das bases de dados do CEFET-ES .....	111
Figura 37 – Resultados de classificação com exemplares de dados da 1ª fase ....	113
Figura 38 – Resultados de classificação com exemplares da 2ª fase x 1ª fase .....	114
Figura 39 – Sistema de identificação: curvas para a fusão com as taxas referentes à 1ª fase .....	116
Figura 40 – Sistema de verificação: curvas com as taxas de acertos, identidade rejeitadas e identidades aceitas como de outra classe .....	117
Figura 41 – Curvas FAR e FRR: 3000 coeficientes na fusão face-forma de caminhar .....	118
Figura 42 – Curvas FRR e FAR: fusão face-forma de caminhar considerando 5000 coeficientes PoV .....	119

Figura 43 – Curvas FRR e FAR: fusão face-forma de caminhar considerando 4000 coeficientes PoV .....	119
Figura 44 – Curvas FRR e FAR: imagens faciais considerando 4000 coeficientes PoV.....	120
Figura 45 – Curvas FRR e FAR: seqüências do caminhar considerando 4000 coeficientes PoV.....	121

# Lista de siglas

AVI - Audio Video Interleave

BFGS - Broyden, Fletcher, Goldfarb and Shanno

BIC - Bayes Information Criterion

CASIA - Chinese Academy of Sciences - Institute of Automation

CEFET-ES - Centro Federal de Educação Tecnológica do Espírito Santo

CTFD - Continuous Two Frame Difference

DTW - Dynamic Time Warping

EER - Equal Error Rate

ESC - Energia das Silhuetas do Caminhar

FAR - False Acceptance Rate

FRR - False Rejection Rate

FT - Fourier Transform

FWT - Fast Wavelet Transform

GATECH - Georgia Institute of Technology

HMM - Hidden Markov Models

ICA - Independent Component Analysis

K-NN - K-Nearest Neighbor

MAP - Máximo a Posteriori

MEI/MHI - Motion-Energy-Image/Motion-History-Image

NLPR - National Laboratory of Pattern Recognition

ORL - Olivetti Research Laboratory

PCA - Principal Component Analysis

PIN - Personal Identification Number

PoV - Proportion of Variances

RBF - Radial Basis Function

RNA - Redes Neurais Artificiais

STFT - Short Time Fourier Transform

SVD - Singular Value Decomposition

SVM - Support Vector Machines

UCSD - University of California - San Diego

WT - Wavelets Transforms

# Sumário

## Capítulo 1

Introdução.....	19
1.1 Objetivo e motivação para a pesquisa .....	21
1.2 As principais etapas do trabalho .....	21
1.3 A evolução das pesquisas e o estado da arte.....	22
1.3.1 Reconhecimento humano através da forma de caminhar .....	23
1.3.2 O uso de imagens faciais no reconhecimento humano .....	27
1.3.3 Fusão de imagens faciais-forma de caminhar no reconhecimento humano .....	28

## Capítulo 2

O Reconhecimento Humano com o uso da Biometria .....	31
2.1 As taxas de erros e acertos em sistemas biométricos .....	34
2.2 Os vários métodos biométricos atuais em estudo.....	38
2.3 O reconhecimento pessoal através de imagens faciais .....	40
2.3.1 Os módulos de um sistema clássico de reconhecimento facial.....	41
2.4 O caminhar humano usado como um padrão biométrico.....	43
2.4.1 O ciclo do caminhar .....	45
2.4.2 Métodos para obtenção das características do ciclo do caminhar .....	46

## Capítulo 3

Métodos de Extração, Seleção e Classificação de Características .....	48
3.1 Normalização das características biométricas .....	48
3.1.1 Normalização de mínimo-máximo .....	49
3.1.2 Normalização de média zero .....	49

3.1.3 Normalização através da tangente hiperbólica.....	49
3.2 A Técnica da proporção de variâncias .....	50
3.3 A Análise de componentes principais ( <i>PCA–Principal Component Analysis</i> ) .	51
3.4 A análise de componentes independentes ( <i>ICA - Independent Component Analysis</i> ) .....	53
3.5 As transformadas wavelets .....	54
3.6 Classificadores baseados em máquinas de vetores suporte .....	58
3.6.1 Classificador SVM para duas classes linearmente separáveis.....	59
3.6.2 Classificador SVM para duas classes linearmente não-separáveis .....	61
3.7 O classificador baseado em redes neurais .....	62
3.7.1 O classificador neural Bayesiano .....	63
3.7.2 O classificador neural por redes de base radial.....	64

## Capítulo 4

Sistemas de Identificação Pessoal por meio da Forma de Caminhar.....	67
4.1 Extração de dados da sequência de vídeo .....	68
4.2 A detecção da posição de cada silhueta na sequência de vídeo .....	68
4.3 A silhueta obtida via subtração do fundo .....	69
4.4 A silhueta externa e o vetor de características .....	70
4.5 Implementação com abordagem Wavelets-PCA e classificador neural .....	71
4.5.1 Os dados e o procedimento experimental .....	72
4.6 Implementação com abordagem wavelets e classificador SVM .....	75
4.6.1 A base de dados e o procedimento experimental.....	77
4.7 Implementação com abordagem PoV-ICA e classificador SVM .....	81
4.7.1 A extração e seleção de características .....	82
4.7.2 O treinamento e teste dos classificadores .....	84
4.8 Implementação com abordagem PoV e classificador RBF .....	86



4.8.1 A Base de Dados utilizada .....	88
4.8.2 As Etapas de pré-processamento .....	88
4.8.3 A Energia das silhuetas do caminhar .....	89
4.8.4 Os Classificadores utilizados: Euclidiano e RBF .....	90
4.9 Considerações finais.....	91

## Capítulo 5

A Identificação Pessoal por meio da Fusão Face-Forma de Caminhar.....	94
5.1 As diversas formas de se estabelecer a fusão biométrica .....	95
5.2 A fusão de sistemas biométricos com múltiplas fontes de informações.....	95
5.3 As propostas para os esquemas de fusão a serem utilizadas .....	97
5.3.1 As bases de dados de domínio público .....	98
5.3.2 As Etapas de pré-processamento .....	99
5.4 Implementação da fusão no nível de decisão .....	100
5.4.1 A etapa de extração e seleção de características .....	102
5.4.2 O procedimento de classificação.....	102
5.4.3 Análise dos resultados .....	103
5.5 Implementação da fusão no nível das características.....	104
5.5.1 Resultados obtidos com as bases de dados ORL/GATECH .....	105
5.6 Considerações finais.....	108

## Capítulo 6

As Bases de Dados do CEFETES e a Fusão no Nível das Características .....	109
6.1 As etapas de aquisição e as características das bases de dados .....	110
6.2 Resultados considerando o sistema no modo de identificação.....	112
6.2.1 Resultados obtidos com os dados da 1ª fase.....	112
6.2.2 Resultados obtidos com os dados da 2ª fase em relação à 1ª fase .....	113

6.3 Análise das taxas de acertos em função dos coeficientes PoV .....	115
6.4 Análise das taxas de falsa aceitação e falsa rejeição .....	117
6.5 Considerações finais.....	121
Capítulo 7	
Conclusão.....	124
7.1 Sumário de contribuições.....	125
7.2 Pesquisas futuras.....	126
REFERÊNCIAS .....	128

# Capítulo 1

## Introdução

O mundo atual, diante dos avanços da tecnologia da informação, da necessidade de segurança de bens materiais e das pessoas, e do grande número de transações que são realizadas diariamente entre organizações e indivíduos, através de dispositivos eletrônicos inter-conectados, exige cada vez mais o estabelecimento automático da identidade das pessoas por técnicas biométricas que se baseiam na utilização das características fisiológicas e/ou comportamentais.

Atualmente, existe uma boa quantidade de técnicas biométricas que são largamente utilizadas ou que ainda se encontram sob investigação de pesquisadores, dentre as quais podemos citar aquelas baseadas em: impressões digitais, geometria das mãos, veias das mãos, assinatura, íris, padrão da retina, timbre de voz, termogramas faciais, imagem da face e, mais recentemente, a forma de caminhar. Cada uma dessas técnicas tem suas vantagens e desvantagens, e são admissíveis ou não em função das aplicações a que são submetidas.

Os métodos atuais de reconhecimento pessoal baseados em imagens, tais como impressões digitais, faces, íris e retina, ou outras modalidades biométricas, como o timbre de voz, geralmente requerem a cooperação da pessoa para permitir que o sistema possa capturar as informações e características mais importantes, implicando em uma proximidade física e, às vezes, até no contato do indivíduo com o sensor. Ou seja, estes métodos não podem, de forma confiável, reconhecer indivíduos não-cooperativos, à distância e sob a mudança de condições ambientais ou, ainda, em alguns casos, em ambientes com baixo nível de iluminação ou escuros.

O reconhecer automático de indivíduos, pela forma como caminham, é um método biométrico e comportamental relativamente novo, além disso, ele possibilita o reconhecimento à distância e tem a capacidade de funcionar, se forem usadas câmaras infravermelhas ou termográficas, em ambientes sem iluminação. No entanto, ele apresenta algumas limitações, pois pode ser afetado pelo tamanho e formato das vestes, pelo uso de sapatos, de bolsas ou pelo contexto ambiental.

Além disso, as condições físicas da pessoa, como lesões causadas por acidentes ou doenças adquiridas, podem também alterar o seu estilo de andar. Neste caso, se vier a ocorrer uma grande variação no padrão de caminhar da pessoa sob circunstâncias diversas, seja de forma intencional ou involuntária, então o poder discriminatório do sistema de reconhecimento automático irá diminuir. Entretanto, em situações normais, o caminhar humano é inerentemente característico de um indivíduo, tornando-o útil em aplicações de segurança (HAN; BHANU, 2006).

Os sistemas automáticos de reconhecimento pessoal por métodos biométricos ou comportamentais, em geral, envolvem áreas de estudos distintas, como processamento digital de sinais e de imagens, visão computacional, inteligência artificial e reconhecimento de padrões, entre outras. Tais ferramentas permitem extrair, selecionar e classificar os padrões obtidos das mais diversas características biométricas, como impressões digitais, voz, íris, retina, assinatura, imagens faciais e de seqüências de vídeo do caminhar.

Pesquisas recentes aplicadas ao reconhecimento pessoal automático propõem a utilização de características biométricos e comportamentais extraídas de imagens infravermelhas ou termográficas aplicadas ao reconhecimento pessoal (HAN; BHANU, 2005); (SOCOLINSKY, et al., 2001). Além disso, alguns estudos em sistemas biométricos multi-modais aplicados ao reconhecimento pessoal automático procuram utilizar ferramentas que propiciem a fusão das informações em diversos níveis de um conjunto de dados biométricos e comportamentais (JAIN; ROSS, 2004); (SINGH, et al., 2004).

Neste trabalho, propomos uma forma de fusão de informações no nível das características para um sistema biométrico bi-modal de reconhecimento pessoal. As características biométricas e comportamentais consideradas na validação deste

trabalho são obtidas de imagens faciais e silhuetas extraídas das seqüências de vídeo da forma humana de caminhar.

### **1.1 Objetivo e motivação para a pesquisa**

Apesar da maioria dos sistemas automáticos de reconhecimento humano por características biométricas apresentarem altas taxas de desempenho, de uma forma geral, eles apresentam algumas desvantagens, tais como a exigência mais intrusiva ou invasiva na obtenção das informações pelos sensores, além de exigirem a colaboração do indivíduo e serem incapazes de efetuar o reconhecimento à distância. Além disso, a capacidade de visão noturna, um componente muito importante em segurança, usualmente não é possível com outros sistemas biométricos. Neste caso, uma câmara de infra-vermelho além de revelar a presença de pessoas, permite extrair silhuetas da forma de caminhar e os detalhes faciais extraídos a partir de imagens capturadas por câmaras localizadas a médias distâncias.

Em consequência disso e diante dos poucos trabalhos disponíveis na literatura, nossa pesquisa sugere arquiteturas e algoritmos que permitem a implementação de um sistema biométrico de reconhecimento e autenticação de indivíduos através da fusão de características faciais e da forma humana de caminhar,

O objetivo principal é o de propor uma estratégia de fusão das informações no nível das características entre as imagens faciais e as seqüências de vídeo da forma de caminhar e avaliar o desempenho do sistema de classificação. Nossa hipótese é que tal arquitetura e algoritmos a serem pesquisados deverão proporcionar um melhor desempenho quando comparado com aqueles obtidos a partir das características individuais formadas pelas imagens faciais ou por seqüências de vídeo do caminhar.

### **1.2 As principais etapas do trabalho**

Em uma primeira etapa, nossa proposta é a de estudar e avaliar os sistemas de reconhecimento pessoal baseados na forma humana de caminhar. Isto porque, quando se tenta identificar uma pessoa em ambientes naturais, como aqueles que se apresentam em aplicações de segurança, algumas características biométricas

como impressão digital, assinatura, voz ou íris não são bem adequadas, devido ao aspecto intrusivo e colaborativo do indivíduo.

Na etapa posterior, diante das vantagens e desvantagens apresentadas para as diversas categorias biométricas por Jain, Ross e Prabhakar (2004), em sistemas de reconhecimento mono-modos, e levando em consideração a característica pouco intrusiva de sistemas faciais, propomos um sistema bi-modal de reconhecimento que utiliza a fusão das características de imagens faciais com as silhuetas extraídas das seqüências de vídeo da forma humana de caminhar.

Para tanto, em uma primeira fase, usamos duas bases de dados de domínio público distintas, uma delas contendo as imagens faciais e a outra, as seqüências de vídeo da forma de caminhar. Neste caso, pelo fato das seqüências de vídeo e as imagens faciais não serem das mesmas pessoas, consideramos que não há correlação entre as duas características e assinalamos as identidades de cada um tendo o cuidado de associar as duas características de forma fixa para os mesmos indivíduos.

Na fase final da pesquisa, objetivando a validação da nossa proposta, bem como alguns aspectos relacionados com a robustez do sistema de reconhecimento, obtivemos duas bases de dados com um total de 58 indivíduos. A base de dados das seqüências do caminhar, para um mesmo indivíduo, foi obtida simultaneamente com a base de dados das imagens faciais. O cenário foi montado em um laboratório do CEFET-ES, onde as capturas das seqüências e imagens ocorreram em duas fases e em épocas diferentes.

### **1.3 A evolução das pesquisas e o estado da arte**

O reconhecimento humano à distância de um caminhante, de acordo com os resultados das pesquisas iniciais em áreas médicas e comportamentais, leva em consideração todos os aspectos de natureza estática, cinemática e dinâmica do corpo em movimento. Sendo assim, para agregarem a maior quantidade de informações possíveis, os sistemas de reconhecimento pessoais baseados em imagens faciais e na forma humana de caminhar devem levar em conta as obtenções das seqüências de vídeo do caminhar e das imagens faciais em ângulos arbitrários com respeito à câmara. Neste caso, o ideal seria o uso de diversas

câmaras para permitir a obtenção do modelo 3-D da pessoa (SHAKHNAROVICH; LEE; DARREL, 2001).

Entretanto, devido aos custos computacionais envolvidos, bem como à possibilidade de operação em tempo real, a solução mais adequada, hoje, e que apresenta bons resultados no reconhecimento automático de uma pessoa através da sua forma de caminhar, se dá quando ela caminha de perfil diante da câmara e ao longo de um cenário ligeiramente controlado. De forma semelhante, no reconhecimento automático facial também é desejável que se obtenham vistas frontais de sua face.

É importante ressaltar ainda trabalhos mais recentes propondo a utilização de imagens infravermelhas e termográficas aplicadas ao reconhecimento pessoal que podem também ser aplicados a imagens extraídas de seqüências de vídeo, tanto para faces quanto para a forma de caminhar (SINGH et al., 2004 ) (SOCOLINSKY; SELINGER, 2004) (HAN; BHANU, 2005).

### **1.3.1 Reconhecimento humano através da forma de caminhar**

A crença de que os seres humanos podem distinguir entre os diversos padrões da forma de caminhar de diferentes indivíduos é conhecida há muito tempo. Entre as características relacionadas que discriminam a forma de caminhar, incluem-se a largura do passo, o ritmo do caminhar e, provavelmente, outros fatores tais como a postura, o balanço dos braços, dos quadris, ombros e do corpo.

A sugestão de que os seres humanos podem identificar outras pessoas por sua forma de caminhar foi inicialmente investigada na década de 70, em uma série de estudos. A partir dessas pesquisas iniciais, diversos outros trabalhos relacionando a forma de caminhar com a identificação pessoal ocorreram em pequenos passos, conforme apresenta a síntese de evolução histórica:

- Cuttting e Kozlowski (1977) demonstraram que os observadores poderiam, com um certo grau de confiabilidade, reconhecer a si mesmo e a seus amigos a partir das exposições de poses dinâmicas constituídas por pontos de luz. Ele apresentou aos participantes as suas próprias imagens que tinham sido reduzidas a um conjunto de silhuetas constituídas de diversos pontos de luz. Suas experiências sugeriram que nós temos alguma noção implícita do

movimento humano, e podemos reconhecer dados temporais dentro deste contexto.

- Kozlowski e Cutting (1977) investigaram primeiramente se os observadores poderiam identificar o sexo de um caminhante com sua silhueta representada por pontos de luz. Os resultados indicaram uma taxa de acertos de 65% a 70% quando as silhuetas do caminhante eram vistas de perfil, sugerindo, pela primeira vez, o caráter biométrico-comportamental da forma humana de caminhar.
- Barclay, Cutting e Kozlowski (1978) sugeriram que os estilos individuais da forma de caminhar podem ser capturados por análise das diferenças encontradas em uma série básica de movimentos pendulares dos membros, dando os primeiros passos para as pesquisas da forma de caminhar com abordagens baseadas em modelos para uso no reconhecimento pessoal.
- Cutting e Proffitt (1981), em sua pesquisa, demonstram que o sexo de um indivíduo caminhante pode ser identificado indiretamente com a determinação do "centro de gravidade" da sua silhueta. Essa pesquisa teve uma importância fundamental para os futuros trabalhos em reconhecimento pessoal através da forma de caminhar com base em abordagens holísticas, por fornecer a introspecção necessária aos pesquisadores de como os observadores puderam discriminar entre padrões do caminhar de diferentes indivíduos.
- Dittrich (1993), fazendo uso da técnica dos pontos de luz, apresentou resultados ainda mais promissores. Seus estudos demonstraram que, não somente era possível uma pessoa andando rapidamente ter sua silhueta extraída dos pontos de luz, mas que, também, um observador dessas silhuetas poderia distinguir entre diferentes movimentos, incluindo o caminhar, saltar e subir escadas.
- Stevenage, Nixon e Vince (1999) exploraram a habilidade das pessoas em identificar outras usando tão somente as informações do caminhar. Eles perceberam que, mesmo quando os observadores não eram familiarizados



com as pessoas caminhantes, diante de um breve tempo de exposição, eles podiam identificá-los corretamente com uma grande taxa de acertos.

- Schollhorn et al (2002), em um estudo mais recente, analisaram a forma de caminhar de quinze pessoas, cujo objetivo era avaliar e detectar a presença da identidade de informações ou assinaturas associadas aos padrões necessários à identificação. Neste estudo, eles encontraram que as variáveis cinéticas capturadas, usando-se uma plataforma de medição de força, bem como as variáveis cinemáticas capturadas por marcadores reflexivos na coxa, na perna e no quadril, eram ambas necessárias para a identificação do indivíduo via forma de caminhar. Além disso, uma grande contribuição do estudo é que ele mostrou que simplesmente o uso de algumas partes da perna e do corpo era suficiente para uma boa identificação.

As abordagens utilizadas no problema do reconhecimento pessoal pela forma de caminhar, em geral, podem ser baseadas em modelos ou podem ser livres de modelos, também chamadas de representações holísticas (BOULGOURIS; HATZINAKOS; PLATANIOTIS, 2005). Ambas seguem os procedimentos de extração, seleção e correspondência de características associadas com a classificação, e exigem um alto nível de processamento. A maior diferença entre elas diz respeito à forma com que se estabelece a correspondência entre dois quadros consecutivos.

Os métodos que assumem um modelo à priori procuram por correspondências de características entre seqüências de imagens em 2-D com fenômenos físicos baseados em movimentos periódicos, que permitem modelar os dados por um conjunto de equações. Os dois exemplos mais importantes dessa abordagem incluem os trabalhos de:

- Cunado et al.(1999a), que procuram modelar o ângulo de rotação do quadril como um pêndulo simples, onde consideram um modelo para a forma de caminhar como um movimento harmônico simples, e evoluem para a extração das características da forma de caminhar pelo ajuste dos ângulos e movimentos das pernas ao modelo de um pêndulo articulado.

- Lee e Grimson (2002), seguindo a linha do ajuste de figuras geométricas a determinadas partes do corpo, apresentam uma proposta onde várias elipses são ajustadas a diversas partes das silhuetas representando o corpo em movimento. Nesse caso, os parâmetros dessas elipses, tais como o centróide e sua excentricidade, são usados para representar as características da forma de caminhar da pessoa, sendo que o reconhecimento é atingido através da comparação dos modelos.

Os métodos estatísticos, livres de um modelo ou abordagens holísticas, por sua vez, procuram estabelecer, mesmo que implicitamente em seus vetores de características, correspondências entre quadros sucessivos baseando-se, principalmente, em predição ou estimação de características relacionadas com a posição, velocidade e forma assumida pelas silhuetas nas seqüências de vídeo. Alguns exemplos dessa abordagem incluem, entre outros, os trabalhos de:

- Little e Boyd (1998), os quais extraíram os vetores de características das imagens em movimento a partir da densidade do fluxo ótico e usaram os momentos relacionados à frequência e fase estimados das seqüências para estabelecerem o reconhecimento pessoal por sua forma de caminhar. Esse trabalho tem uma grande importância para o surgimento de diversas outras técnicas, principalmente pela liberação dos autores para domínio público da base de dados UCSD, que é, inclusive, utilizada na validação de algumas das nossas propostas.
- Huang, Harris e Nixon (1998), os quais consideram as seqüências de imagens das silhuetas de um ciclo do caminhar para compor o vetor de características básico. A análise de componentes principais é então aplicada, dando origem às chamadas “eigengaits” que são, em seguida, usadas por um classificador.
- Benalbelkader, Davis e Cutler (2002), os quais usam desenhos de auto-similaridade de imagens para obter a correlação entre duas silhuetas de uma seqüência. Em seguida, utilizam a técnica PCA (*Principal Component Analysis*) para redução de dimensionalidade e o classificador K-vizinhos mais próximos para avaliar o desempenho do sistema.

- Kale et al. (2002a), os quais usaram um conjunto de vetores extraídos das silhuetas de seqüências de vídeo para constituir os estados de um modelo baseado em HMM (*Hidden Markov Models*) contínuo.
- Boulgouris, Hatzinakos e Plataniotis (2005), que utilizaram um algoritmo baseado na deformação dinâmica do tempo (DTW - *Dynamic Time Warping*) para o reconhecimento da forma de caminhar. O algoritmo compara duas seqüências de silhuetas ao computar a distância entre duas silhuetas como uma função do tempo.

### 1.3.2 O uso de imagens faciais no reconhecimento humano

Os primeiros trabalhos propostos para o reconhecimento pessoal através de imagens faciais foram baseados na aplicação da análise dos componentes principais (PCA – *Principal Component Analysis*) ou “eigenfaces”, sendo o mais importante, e que serviu de referência para vários outros, o de Turk e Pentland (1991). Desde a sua publicação, as abordagens estatísticas para se obter as características do modelo facial são as mais empregadas.

Neste caso, cada pose bidimensional da imagem, cujos pixels detêm as características da mesma, é tratada como um vetor obtido através da varredura da imagem em ordem lexicográfica, em que a dimensão do vetor é dada pelo número dos pixels da imagem. Isto é, dada uma imagem  $I_q[N \times M]$  de um conjunto  $Q$  de exemplares cujos pixels são representados por  $p_{n,m}$ , onde  $n = 1, 2, 3 \dots N$  e  $m = 1, 2 \dots M$ , então o conjunto de todos os pixels da imagem escritos em ordem lexicográfica será

$$P_q = (p_{1,1}, p_{1,2}, \dots, p_{1,M}, p_{2,1}, p_{2,2}, \dots, p_{2,M}, \dots, p_{N,1}, p_{N,2}, \dots, p_{N,M}), \quad (1)$$

onde  $P_q^T$  é um vetor coluna de dimensão  $N \times M$ . O conjunto transposto de vetores-imagem  $P$  obtidos dos  $Q$  exemplares é então concatenado para formar uma matriz de faces  $F[N \times M, Q]$  a qual pode ser aplicada as mais diversas formas de seleção de características.

Em seguida, após todos os vetores correspondentes a cada imagem serem concatenados para formar a matriz  $F[N \times M, Q]$ , Turk e Pentland (1991) aplicaram o

PCA sobre a matriz de covariância associada à respectiva matriz de faces para gerar um subespaço linear representado pelos seus auto-vetores.

Tipicamente, o número de auto-vetores que discriminam as faces é muito menor que a dimensão do espaço original. O reconhecimento consiste, então, em encontrar naquele subespaço a face que mais se aproxima daquela apresentada.

Vários trabalhos importantes aplicando as mais diversas técnicas contribuíram para elevar o nível das pesquisas com o uso de imagens faciais em sistemas de reconhecimento pessoal. Zhao et al. (2003) apresentam um artigo detalhando e criticando todo esse processo evolutivo, incluindo os trabalhos desenvolvidos até a data de sua publicação, em reconhecimento facial.

- **Reconhecimento facial a partir de seqüências de Vídeo**

Algumas abordagens para executar o reconhecimento facial baseadas nas seqüências de vídeo foram também propostas. Entretanto, a maioria delas faz uso de um pequeno conjunto de imagens extraídas da seqüência de quadros. Entre as mais relevantes, destacamos:

- Choudhury et al. (1999), que detectam e capturam os frames da face em uma seqüência de vídeo e, em seguida, executam o reconhecimento baseando-se em uma daquelas imagens que satisfazem a determinados critérios em relação a uma pose padrão.
- Biuk e Loncaric (2001), os quais fazem o reconhecimento através da aplicação de PCA sobre um vetor característico extraído da seqüência de vídeo da face de uma pessoa, que apresenta bons resultados mesmo ao considerar-se uma rotação da cabeça em um ângulo de aproximadamente 180°.

### **1.3.3 Fusão de imagens faciais-forma de caminhar no reconhecimento humano**

Com a necessidade de confiabilidade na autenticação pessoal, o uso dos sistemas automáticos de verificação da identidade baseados em características biométricas múltiplas está cada vez mais difundido. Como exemplos de sistemas que apresentaram bom desempenho, temos a integração de imagens faciais e

impressões digitais (HONG; JAIN, 1998), a fusão de imagens faciais e voz, que apresentou 100% de desempenho (CHOUDHURY et al.,1999), e a combinação de imagens faciais com imagens da íris (WANG; TAN; JAIN, 2003).

Os sistemas de reconhecimento faciais e da forma de caminhar são hoje muito pesquisados por serem pouco ou nada intrusivos, e permitirem a identificação à distância, apesar de não apresentarem uma alta precisão, como ocorre com outros sistemas biométricos mais intrusivos, ou até invasivos, e que requerem a colaboração do indivíduo como é o caso das impressões digitais, retina e íris.

O problema da baixa precisão na verificação da identidade pessoal por faces ou forma de caminhar, que ocorre para sistemas uni-modais, pode ser minimizado, fundindo-se os dois sistemas biométricos de identificação. Assim, tal como comentado anteriormente, além de aumentar a cobertura populacional atendida em consequência de alguma lesão que não permita a atuação de um sistema isolado, é sabido ainda que, de uma forma geral, a fusão de distintas características biométricas permite diminuir a taxa de erro total, dada pelas taxas de falsa aceitação e de falsa rejeição (HONG; JAIN, 1998).

Além disso, o fato inconveniente presente em outras modalidades onde, para se efetuar a fusão delas, o usuário deve se submeter a vários estágios de aquisição, a combinação face e forma de caminhar permite a aquisição simultânea de imagens isoladas ou em forma de seqüências de vídeo. Portanto, neste caso particular, nenhuma inconveniência adicional é introduzida. Alguns exemplos de trabalhos que mais se destacam nesta área são os de:

- Shakhnarovich, Lee e Darrel (2001) que, a partir do uso de várias câmaras, obtêm um modelo tridimensional para o corpo, formado pelas silhuetas e, para a face, formado pelas imagens faciais, conseguindo 91% como taxa máxima de acertos.
- Kale, Roy-Chowdhury e Chellappa (2002b), que fizeram uso das silhuetas e estabeleceram a fusão com as imagens faciais utilizando-se das abordagens hierárquica e holística, alcançando resultados com taxas de acertos variando de 93% a 97% para a primeira e 100% de acertos para a segunda abordagem.

No próximo capítulo serão apresentados, de forma sucinta, algumas técnicas utilizadas sobre as características extraídas dos indicadores biométricos que viabilizam a implementação de sistemas automáticos de reconhecimento humano. Serão também apresentadas comparações entre alguns indicadores biométricos dando uma maior relevância, devido à proposta deste trabalho, aos sistemas baseados em seqüências de vídeo da forma de caminhar e nas imagens faciais.

# Capítulo 2

## O Reconhecimento Humano com o uso da Biometria

Um sistema de identificação biométrica é, essencialmente, um sistema de reconhecimento de padrões. A Figura 1 apresenta, de uma forma genérica, o diagrama de blocos simplificado de um sistema capaz de desempenhar as funções necessárias para estabelecer a identidade de um indivíduo por comparar a representação das suas características biométricas e comportamentais com os padrões armazenados em uma base de dados (JAIN; ROSS; PRABHAKAR, 2004).

Neste caso, o diagrama consiste de dois módulos, um que detecta e registra o conjunto de informações biométricas e outro que opera na identificação pessoal.

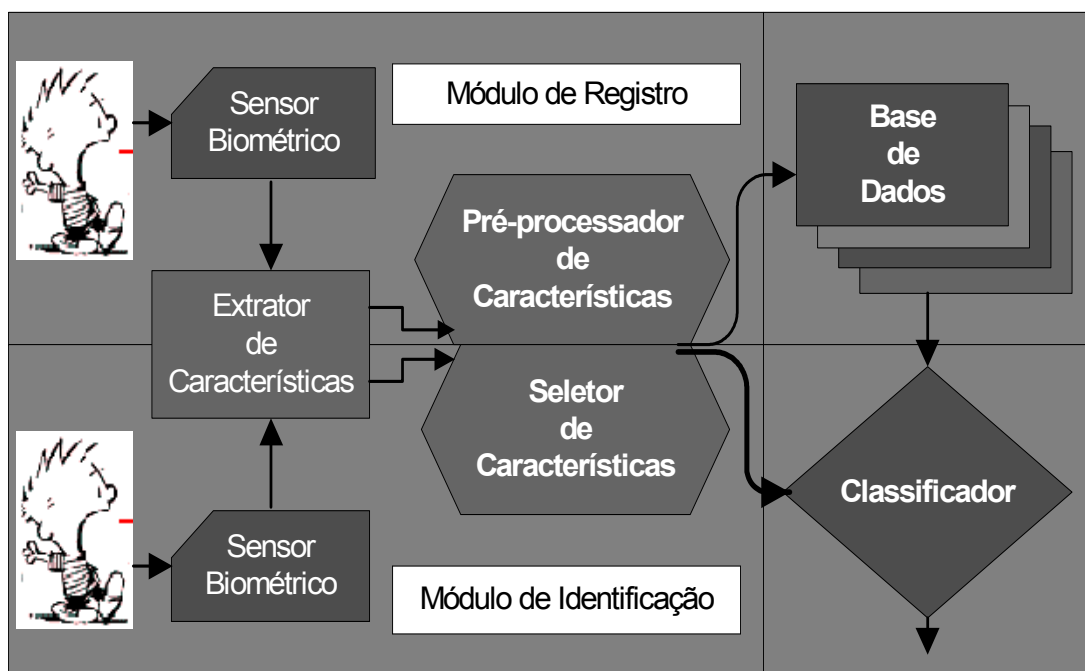


Figura 1 – Arquitetura de um sistema de identificação biométrica genérico

O primeiro módulo adquire os dados biométricos pessoais através de um ou mais sensores, extraíndo, selecionando e indexando um conjunto de características desses dados e armazenando-os em uma base de dados com as classes correspondentes. O segundo módulo estabelece a identificação pessoal comparando e classificando as características biométricas de uma pessoa a ser identificada com o conjunto de características do modelo armazenado na base de dados.

O sistema de reconhecimento pessoal automático por características biométricas consiste em armazenar em uma base de dados de treinamento um determinado número de exemplares daquelas características para cada pessoa associada, bem como uma informação da sua respectiva identidade. Elas podem ser representadas como uma simples imagem ou como um conjunto delas, como uma seqüência de vídeo, ou como um conjunto de poses discretas. Assim, dada uma imagem ou um conjunto de imagens teste, o problema consiste em identificar um indivíduo como membro da base de dados de treinamento através das características usadas no algoritmo de reconhecimento. Portanto, supondo que a base de dados  $BD$  contém um número de pessoas  $M$ , podemos escrever:

$$BD = \{(I_k, ID_k) : k = 1, 2, \dots, M\} \quad (2)$$

onde  $I_k$  é a imagem da pessoa  $k$ , e  $ID_k$  é a identidade associada à referida pessoa.

O desafio é projetar um sistema de reconhecimento que, para toda imagem de entrada, identifique como um dos indivíduos conhecidos ou não, validando uma identidade ou rejeitando-a. Neste projeto, há duas partes principais relacionadas ao problema: a extração e seleção de características e a classificação das identidades.

Na extração e seleção das características, a tarefa é encontrar uma forma eficiente de representar os dados da imagem através dos “pixels” da mesma. Isto envolve encontrar uma função capaz de projetar os vetores de características das imagens do espaço original em um outro de menor dimensão. A extração e seleção de características são aplicadas em todas as imagens da base de dados original, gerando uma outra base de dados  $BD_x$ , geralmente menor que a original, constituída por um conjunto de vetores de características, que pode ser expressa por:



$$BD_x = \{(x_k, ID_k) : k=1, 2, \dots, M\} \quad (3)$$

onde  $x$  é o vetor de características derivado da imagem  $I_k$  e  $ID_k$  é a identidade pessoal associada ao referido vetor.

O problema da classificação, por outro lado, envolve projetar um algoritmo cuja função seja capaz de associar os vetores de características ao rótulo de identidade apropriado da classe à qual pertence, isto é, identificar a pessoa. Como mencionado anteriormente, além das etiquetas da classe das  $M$  pessoas conhecidas  $ID_k$ , o algoritmo deve ser capaz de rejeitar a identidade de pessoas cujos vetores de características não pertencem à base de dados de treinamento. Assim, o problema central do reconhecimento pessoal, por qualquer uma das características biométricas e das técnicas de reconhecimento de padrões utilizadas, é encontrar uma solução otimizada capaz de reconhecer como pessoas genuínas aquelas da base de dados e rejeitar pessoas estranhas à mesma. Além disso, dependendo do contexto da aplicação, o sistema pode operar no modo de verificação ou no modo de identificação (JAIN; ROSS; PRABHAKAR, 2004).

No modo de verificação, o sistema valida a identidade apresentada pela pessoa através da comparação do dado biométrico capturado com seus próprios modelos armazenados na base de dados. Isto é, o indivíduo que deseja ser reconhecido apresenta uma informação de sua própria identidade (*PIN, User name, Smart card*) que será validada ou não. Neste caso, tal problema pode ser formulado como segue: dado como entrada um vetor de características  $X_Q$  e uma identidade  $I$ , e como saída um conjunto de modelos biométricos pessoais  $X_I$  armazenados na base de dados, o sistema deve determinar se  $(I, X_Q)$  são pertencentes às classes  $w_1$  de usuário genuíno ou  $w_2$  de um impostor. Assim, o sistema deve ser capaz de verificar a seguinte assertiva:

$$(I, X_Q) \in \begin{cases} w_1, & \text{se } S(X_Q, X_I) \geq l \\ w_2, & \text{caso contrário} \end{cases} \quad (4)$$

onde  $S$  é a função de similaridade entre  $X_Q$  e  $X_I$ , e  $l$  é um limiar de decisão pré-definido.

No modo de identificação, o sistema identifica uma pessoa pela comparação do dado biométrico capturado com todos os modelos armazenados na base de dados em um relacionamento do tipo um para vários, possibilitando o indivíduo ser reconhecido ou não sem apresentar qualquer informação de sua própria identidade. Neste caso, tal problema pode ser formulado como segue: dado como entrada um vetor de características  $X_Q$ , um conjunto de identidades genuínas  $I_k$ ,  $k \in \{1, 2, 3, \dots, N\}$  e outra identidade  $I_{N+1}$ , que deve ser rejeitada quando a solicitação ao sistema for feita por um impostor, e como saída um conjunto de modelos biométricos pessoais  $X_{Ik}$  armazenados na base de dados, o sistema deve determinar se  $X_Q$  é pertencente a  $I_k$  ou a  $I_{N+1}$ , conforme expresso na seguinte assertiva:

$$X_Q \in \begin{cases} I_k, & \text{se } \max_k \{S(X_Q, X_{Ik})\} \geq l \quad k=1, 2, \dots, N \\ I_{N+1}, & \text{caso contrário} \end{cases} \quad (5)$$

onde  $S$  é a função de similaridade entre  $X_Q$  e  $X_{Ik}$ , e  $l$  é um limiar de decisão pré-definido.

## 2.1 As taxas de erros e acertos em sistemas biométricos

Em sistemas de reconhecimento pessoal, duas amostras da mesma característica biométrica nunca são exatamente iguais. Isso ocorre devido a ruídos no sensor, alterações fisiológicas ou comportamentais ou, ainda, devido a interação do indivíduo com os sensores e as modificações no ambiente em que se verifica a coleta. Portanto, devido às variações intra-classes nas características biométricas, a identificação pode ser estabelecida somente com determinado grau de confiabilidade.

A decisão a ser tomada se encaixa em uma de duas situações: ou se tem a identidade verdadeira de um indivíduo, ou a identidade de um impostor. Para cada uma delas, existem duas possibilidades a serem retornadas pela função de similaridade do classificador: falsa ou verdadeira. Sendo assim, tem-se um total de quatro possibilidades:

a - um indivíduo genuíno tem sua identidade verificada e é aceito,

b – um indivíduo genuíno tem sua identidade não verificada e é rejeitado,

c - um impostor não é reconhecido e é rejeitado; e

d - um impostor tem uma identidade verificada e é aceito.

Está claro, portanto, que os rótulos (a) e (c) são decisões corretas do sistema de classificação.

A confiabilidade do sistema de reconhecimento ao ser associada com as decisões (b) e (d) pode ser utilizada para estabelecer duas taxas de erros:

- a Taxa de falsa aceitação (*FAR - False Acceptance Rate*), que representa a situação em que impostores são aceitos como indivíduos genuínos e,
- a Taxa de falsa rejeição (*FRR - False Rejection Rate*), que ocorre quando indivíduos genuínos são rejeitados como se fossem impostores.

Em geral, os classificadores dos sistemas de reconhecimento pessoal usam escores para expressarem a similaridade entre um padrão de teste e um modelo biométrico. Como descrito anteriormente, no modo de identificação, o acesso ao sistema é concedido somente se o escore para a pessoa presente na base de dados de treinamento superar um valor limiar de decisão. Por outro lado, no modo de verificação, o padrão teste e a identidade a serem verificados são comparados ao modelo da base de dados e se o classificador emitir um escore superior a um determinado limiar e houver coincidência da identidade apresentada, então o indivíduo será aceito como genuíno.

Na teoria, os escores dos padrões de teste das pessoas conhecidas pelo sistema devem sempre ser mais elevados do que os escores dos impostores. Tomando este fato como verdadeiro, um único ponto inicial que separa os dois grupos de escores, poderia ser usado para diferir entre indivíduos genuínos e impostores. Entretanto, devido a diversas razões, esta suposição não é verdadeira para sistemas biométricos do mundo real, isto é, em alguns casos os padrões de teste dos impostores geram escores maiores do que alguns padrões de teste de indivíduos genuínos; e, por essa razão, aquele ponto inicial escolhido dos escores e tomado como limiar de decisão de classificação, gera alguns erros de identidade.

Sendo assim, se o limiar de escores escolhido for muito elevado, então nenhum ou uma quantidade muito pequena dos escores de impostores irá superá-lo, entretanto, os escores de indivíduos genuínos com valores menores que o limiar escolhido serão rejeitados. Por outro lado, se o limiar de escores escolhido for muito baixo, uma pequena quantidade ou nenhum padrão de teste de indivíduos genuínos serão falsamente rejeitados, porém, alguns padrões de teste de impostores serão falsamente aceitos.

Para elucidar um pouco mais esse dilema, imaginemos um sistema de verificação biométrico contendo os modelos e identidades em uma base de dados treinada. O objetivo aqui é o de testar uma grande quantidade de padrões de indivíduos genuínos e impostores. Neste caso, como apresenta a Figura 2, uma distribuição normal gaussiana irá representar, respectivamente, a frequência de identidades rejeitadas ou aceitas em função dos escores distribuídos em torno de uma determinada média. Como era de se esperar, temos duas distribuições em que a média é maior para a distribuição dos escores atribuídos a indivíduos genuínos em relação aos impostores.

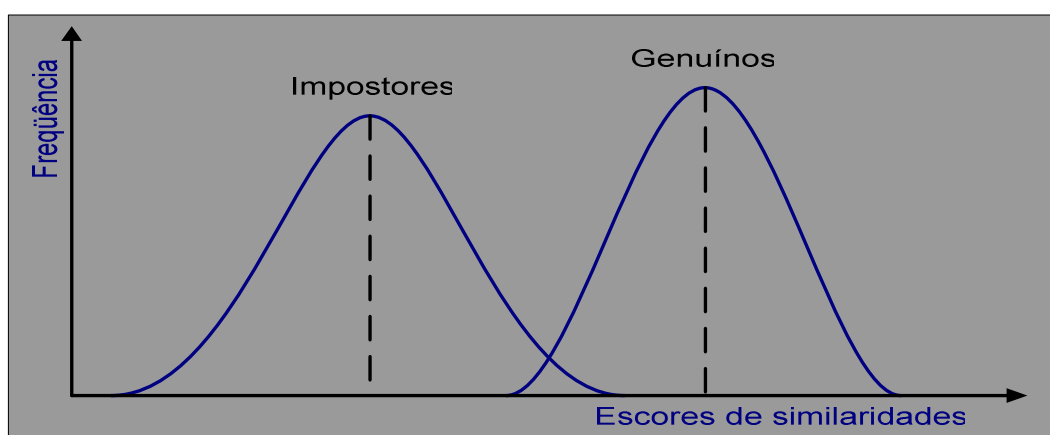


Figura 2 – Curvas representando a frequência de erros e acertos em função dos escores

Assim, em função dos escores que superam o limiar de aceitação, a fração dada pelos padrões de teste de impostores aceitos ( $PIa$ ) dividida pelo número total de padrões testes dos impostores ( $PI$ ) é chamada de taxa de falsa aceitação ( $FAR$  - *False Acceptance Rate*), e seu valor é unitário se todos os padrões testes de impostores forem aceitos, e zero, em caso contrário.

$$FAR = \frac{PIa}{PI} \quad (6)$$

Por outro lado, ainda em função dos escores que superam o limiar de aceitação, a fração dada pelo número de padrões teste de indivíduos genuínos rejeitados ( $PGr$ ) dividido pelo número total de padrões de testes genuínos ( $PG$ ) é chamada de taxa de falsa rejeição ( $FRR$  - *False Rejection Rate*), e seu valor também ficará entre zero e um.

$$FRR = \frac{PGr}{PG} \quad (7)$$

A escolha do valor do limiar de decisão ( $I$ ) de escores emitidos pelo classificador para o sistema de reconhecimento, na prática transforma-se em um grande problema, se as distribuições de limiares de escores de indivíduos genuínos e de impostores se sobrepuserem, conforme sugere a Figura 3. Neste caso, observamos que pequenas FRRs geralmente implicam em grandes FARs e vice-versa. Notamos também que, se há uma sobreposição nas distribuições das curvas de impostores e genuínos em função dos limiares de escores, então as curvas FAR e FRR se interceptam em algum ponto onde suas taxas são idênticas. Este ponto é chamado de taxa de erros iguais ( $EER$  - *Equal Error Rate*) e serve como referência na escolha do limiar ou ponto de corte final do classificador.

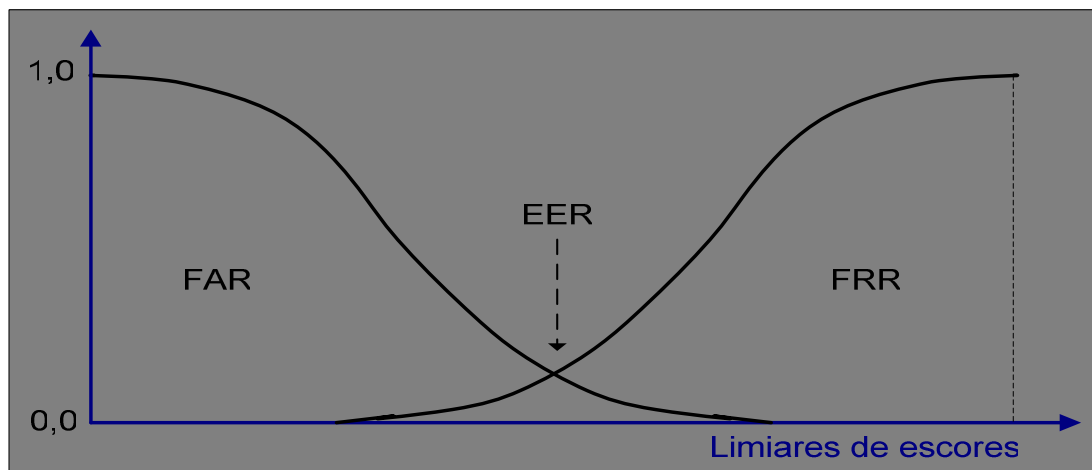


Figura 3 – Curvas representando FAR, FRR e EER

Sendo assim, as duas taxas de falsa aceitação (FAR) e de rejeição (FRR) estão vinculadas a um valor atribuído para o limiar de decisão  $I$ , isto é, se  $I$  decresce para que o sistema seja mais tolerante a ruídos e às variações na entrada, então a FAR aumenta; e se, por outro lado,  $I$  é maior para tornar o sistema mais seguro, então a FRR aumenta. Por estas razões, o desempenho da maioria dos sistemas de

reconhecimento é especificado em termos de uma pequena taxa de falsa aceitação (FAR), uma vez que é mais seguro rejeitar um indivíduo genuíno do que aceitar um impostor (HONG; JAIN, 1998).

## 2.2 Os vários métodos biométricos atuais em estudo

Atualmente, há vários grupos de pesquisa, bem como um grande número de aplicações, que fazem uso de uma ou mais características biométricas em sistemas de reconhecimento pessoal. A Figura 4 apresenta diversas modalidades biométricas em fase de pesquisas e algumas delas já utilizadas em sistemas comerciais.

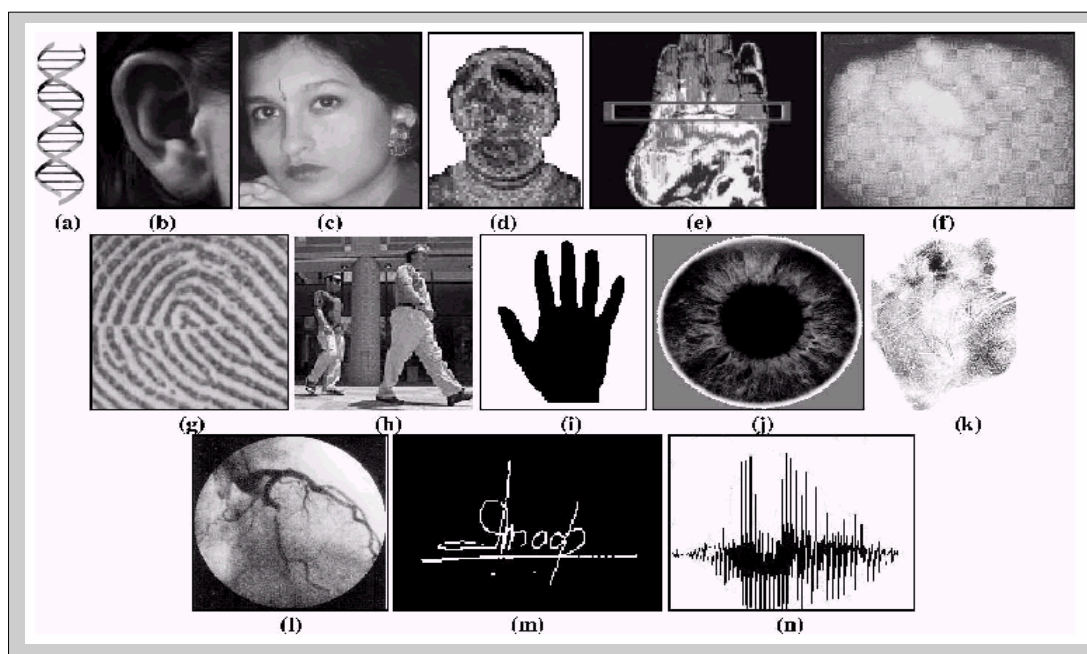


Figura 4 - Exemplos de características biométricas (JAIN; ROSS; PRABHAKAR, 2004)

Neste caso, temos em (a) a estrutura do DNA, (b) orelha, (c) face, (d) termograma facial, (e) termograma da mão, (f) veias da mão, (g) impressões digitais, (h) forma de caminhar, (i) geometria da mão, (j) íris, (k) palma da mão, (l) retina, (m) assinatura, e (n) voz. Além dessas, têm-se ainda o odor e o toque em teclas conforme apresenta a Tabela 1.

Do ponto de vista da aplicação, cada um deles possui as suas vantagens e desvantagens. Assim, nenhum sistema de reconhecimento baseado em uma única característica biométrica é ótimo para qualquer aplicação. Sendo assim, é importante salientar que nenhuma das técnicas irá se sobrepôr às demais, se for considerada

sua aplicabilidade para atender um contexto que deve ser ajustado a qualquer cenário do ambiente.

A Tabela 1 apresenta uma série de fatores associados com os diversos métodos biométricos de reconhecimento humano. Isto é, ela leva em conta fatores que especificam, por exemplo, o quanto a característica biométrica ou comportamental é de aplicação universal, qual o seu poder de distinguibilidade, sua retenção ao longo do tempo, sua facilidade ou não em ser coletada, seu desempenho na classificação, sua aceitabilidade em relação às pessoas e sua circunvenção, que reflete o quanto um sistema biométrico pode ser burlado quando se utiliza métodos fraudulentos.

Tabela 1 - Comparação das várias características biométricas (JAIN; ROSS; PRABHAKAR, 2004).

Alto, Médio e Baixo são denotados por A, M e B, respectivamente.

Identificador Biométrico	Universalidade	Distinguibilidade	Permanência	Coletabilidade	Desempenho	Aceitabilidade	Circunvenção
DNA	A	A	A	B	A	B	B
Orelha	M	M	A	M	M	A	M
Face	<b>A</b>	<b>B</b>	<b>M</b>	<b>A</b>	<b>B</b>	<b>A</b>	<b>A</b>
Termograma facial	A	A	B	A	M	A	B
Impressões digitais	M	A	A	M	A	M	M
Forma de caminhar	<b>M</b>	<b>B</b>	<b>B</b>	<b>A</b>	<b>B</b>	<b>A</b>	<b>M</b>
Geometria das mãos	M	M	M	A	M	M	M
Veias das mãos	M	M	M	M	M	M	B
Imagens da Iris	A	A	A	M	A	B	B
Toque em teclas	B	B	B	M	B	M	M
Odor	A	A	A	B	B	M	B
Palma da mão	M	A	A	M	A	M	M
Retina	A	A	M	B	A	B	B
Assinatura	B	B	B	A	B	A	A
Voz	M	B	B	M	B	A	A

Por ser o foco principal deste trabalho, assinalamos em negrito, na Tabela 1, os fatores biométricos relacionados com os sistemas de reconhecimento pessoal baseados em imagens faciais e em seqüências de vídeo da forma de caminhar. Neste caso, observamos, para ambos os sistemas biométricos, algumas vantagens, tais como: média para alta universalidade, coletabilidade e aceitabilidade; e algumas desvantagens, como: média para baixa permanência, distinguibilidade, desempenho e circunvenção.

Assim, diante desses fatores e dos poucos trabalhos nesta área, direcionamos nossas pesquisas para a implementação e avaliação de sistemas de reconhecimento pessoal que implementam a fusão dessas duas características.

### **2.3 O reconhecimento pessoal através de imagens faciais**

O reconhecimento pessoal automático através de imagens faciais, para fins de controle de acesso e segurança, tem sido pesquisado e aplicado com sucesso ao longo de algumas décadas. Ele não exige contato do indivíduo com o sistema sensor, é automático e muito natural, visto ser este, juntamente com a voz, um dos métodos de reconhecimento que nós, humanos, empregamos todos os dias em nossos encontros pessoais.

A maioria das pesquisas, nesta área, tem focalizado as imagens obtidas no espectro visível. Neste caso, apesar do progresso atingido na última década e dos bons resultados obtidos, alguns problemas continuam ainda sem solução adequada, em consequência, principalmente, das imagens serem formadas pela reflexão da luz, tornando-as difíceis de serem trabalhadas em função da dependência do ângulo de incidência e da potência da fonte luminosa. Além disso, atualmente, os principais desafios estão relacionados com a melhoria do desempenho do sistema de reconhecimento, tendo-se em conta as variações entre as imagens de treinamento e de teste, devido às condições de iluminação, alterações na aparência, alterações devido ao tempo entre a obtenção das imagens e diferentes poses das imagens.

Por outro lado, recentemente, e com resultados promissores, tem ocorrido um aumento considerável das pesquisas no reconhecimento pessoal através de



imagens faciais obtidas a partir de fontes do espectro infravermelho ou de imagens termográficas geradas pela emissão de calor da própria face da pessoa.

Na próxima seção, será descrito um sistema clássico de reconhecimento facial consistindo de todos os módulos necessários na normalização da imagem facial, extração e seleção de características e de classificação.

### 2.3.1 Os módulos de um sistema clássico de reconhecimento facial

Em termos gerais, os dois módulos principais que constituem o sistema de reconhecimento pessoal através de imagens faciais são conectados de forma serial, conforme apresenta o esquema da Figura 5.

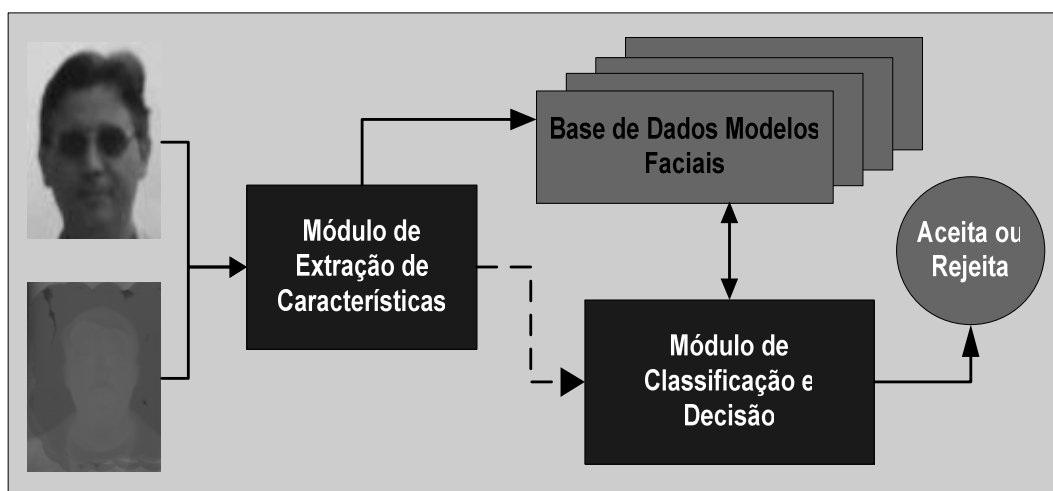


Figura 5 – Sistema clássico de reconhecimento facial

O módulo de extração de características efetua, também, a normalização e seleção das características, e atende tanto à fase de preparação da base de dados para treinamento, quanto à fase de classificação das imagens.

A normalização é responsável pela uniformização de todas as imagens que são usadas para o reconhecimento da face. Isto significa que toda imagem de face, ao ser adicionada à base de dados de treinamento ou utilizada na identificação, tem que ser normalizada primeiramente. A propriedade mais importante do módulo de normalização é o gerenciamento de variações da escala e de rotações verticais da cabeça. As imagens de entrada são escaladas, cortadas e giradas tais que os seus tamanhos, ao serem incorporadas no sistema, sejam exatamente os mesmos, e permitam, conseqüentemente, mais facilidade na comparação entre elas. Uma

característica importante é o corte da imagem, que é um tipo de redução da informação, e consiste em remover o fundo e manter somente as faces ou as partes mais importantes delas, diminuindo o processamento ao remover as informações que não serão úteis ao sistema de reconhecimento.

A segunda função efetuada por este módulo é a extração das características que irão compor o vetor de características básico. Neste caso, independente da imagem ser visível ou infravermelha, o vetor básico de características é obtido através de um algoritmo que varre sequencialmente cada linha da imagem  $N \times M$  e, simultaneamente, concatena os valores de níveis de cinza representados em cada pixel no respectivo vetor que irá ficar com uma dimensão  $D = N.M$ , conforme sugere a Figura 6.

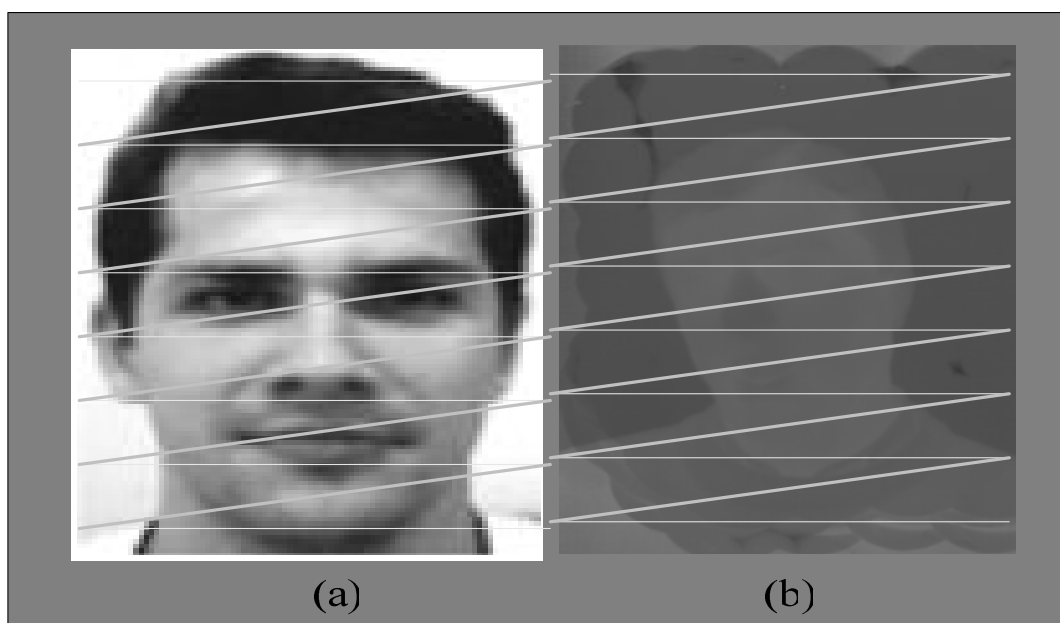


Figura 6 – Obtenção dos vetores de características de imagens faciais

A última função do módulo de extração de características está relacionada com a seleção das características mais discriminatórias de cada vetor básico. Neste caso, diversas técnicas, a serem descritas posteriormente, tais como PoV, PCA, ICA, Wavelets, e etc, podem ser empregadas. Nesta etapa, também pode ocorrer uma redução de dimensionalidade, outra transformação muito importante, para diminuir os custos computacionais e aumentar o desempenho da etapa posterior de classificação.

O módulo de classificação do sistema de reconhecimento facial atribui a cada uma das imagens do candidato na base de dados uma identidade correspondente ou, ainda, informa a ocorrência de uma imagem não identificada pelo sistema. Ele recebe um vetor normalizado do módulo anterior e, durante o reconhecimento, a sua tarefa principal é comparar este vetor correspondente àquela imagem facial de entrada com todos os vetores das imagens armazenadas na base de dados e, em seguida, identificar o candidato, ou, no mínimo, reduzir a quantidade de possíveis candidatos. Uma base de dados grande pode ter muitas imagens armazenadas, tipicamente milhares, sendo que o processamento a ser feito deve produzir resultados rapidamente. Portanto, a comparação de uma imagem teste contra cada uma das imagens treinadas do candidato deve ser muito rápida, mas sem diminuir de forma demasiada a exatidão da resposta. Assim, para todas as imagens na base de dados são obtidos escores baseados em suas distâncias de similaridade em relação à imagem de teste, e os candidatos com os escores mais elevados são escolhidos para o processamento adicional.

## **2.4 O caminhar humano usado como um padrão biométrico**

O corpo humano pode ser definido fisicamente como um complexo sistema de segmentos articulados em equilíbrio estático ou dinâmico, onde o movimento é causado por forças internas que atuam fora do eixo articular, provocando deslocamentos angulares dos segmentos e por forças externas ao corpo. Sendo assim, o caminhar humano é um fenômeno espaço-temporal que caracteriza o movimento de um indivíduo.

A Biomecânica do movimento busca explicar como as formas de movimento dos corpos de seres vivos acontecem na natureza a partir de parâmetros cinemáticos e dinâmicos extraídos de alguma forma desses movimentos. Sendo assim, o caminhar humano, do ponto de vista da biomecânica, consiste em movimentos integrados e sincronizados de centenas de músculos e articulações do corpo, sendo, portanto, um fenômeno espaço-temporal que caracteriza um equilíbrio dinâmico do movimento do indivíduo.

Embora esses movimentos estejam de acordo com o mesmo modelo bipedal básico para todos os seres humanos, eles variam de um indivíduo para outro em certos

detalhes. Assim, de um indivíduo para outro há variações nas temporizações relativas, ângulos formados entre os membros das pernas e magnitudes nas aberturas das mesmas, movimentação dos quadris, do tronco e dos braços, isto é, o caminhar humano difere em muito de uma pessoa para outra.

A forma de caminhar é, presumidamente, única para cada indivíduo simplesmente porque ela é determinada pela totalidade de sua estrutura músculo-esquelética, incluindo, aí, os atributos fisiológicos como massa corporal, comprimento dos membros e estrutura dos ossos. Em pesquisas de análise da forma de caminhar, a caracterização é feita por algumas centenas de parâmetros cinemáticos, como as velocidades e acelerações angulares em determinadas partes e articulações do corpo (BENABDELKADER; DAVIS; CUTLER, 2002).

O atrativo da forma de caminhar como uma técnica biométrica está no fato de que é um método não-intrusivo, podendo ser detectado e medido até mesmo em ambientes com baixa luminosidade e a uma baixa resolução de vídeo. Também é mais difícil para o indivíduo disfarçar essa sua característica do caminhar quando comparado ao aparecimento estático diante de um microfone para captura da voz ou de uma câmara para obter imagens faciais e, ainda mais, essa técnica não requer a cooperação da pessoa (CHELLAPPA; ROY-CHOWDHURY; ZHOU, 2005).

Pesquisas nas áreas médicas e comportamentais sugerem que se todos os movimentos da forma de caminhar pudessem ser representados nos vetores de características, então ela seria única para cada ser humano (NIXON et al., 1999).

Os primeiros trabalhos nesta área que fizeram uso da visão computacional tiveram a preocupação médica em identificar anomalias de natureza biomecânicas, evoluindo para aqueles que classificavam os movimentos em diversas categorias, como caminhar, correr e pular. Mais recentemente, as pesquisas foram direcionadas também para o reconhecimento pessoal (PHILLIPS et al., 2002); (WANG et al., 2003b).

Neste trabalho, foi aplicada a abordagem holística que captura implicitamente as informações temporais envolvidas na seqüência de vídeo. Para tanto, a partir da pessoa caminhando lateralmente diante de uma câmara estática, foram obtidos os vetores de características das silhuetas concatenando-as na ordem em que elas

aparecem na seqüência de vídeo, conforme apresenta o cenário da Figura 7, que sugere uma vista de cima de alguns quadros de uma seqüência do indivíduo caminhando diante da câmara.

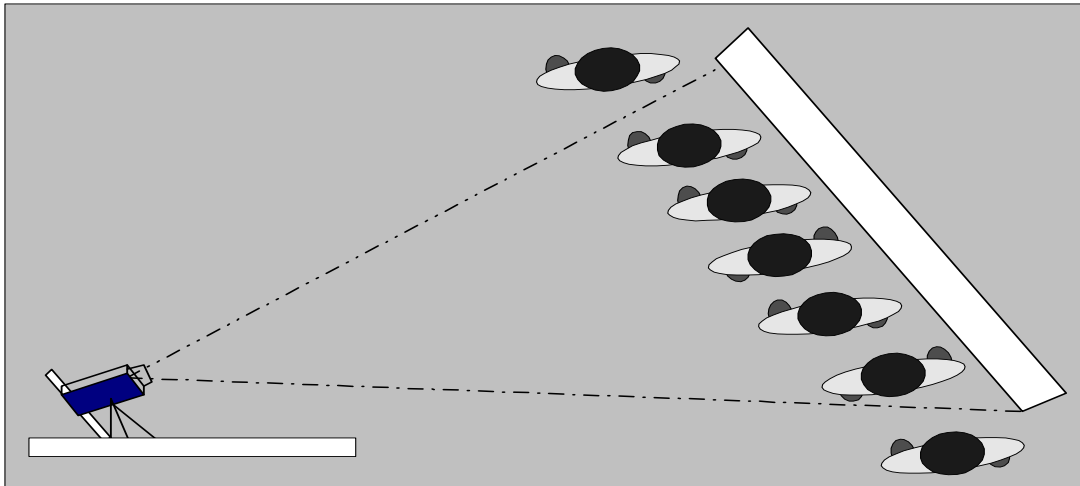


Figura 7 – Vista de cima do cenário para a captura das seqüências de vídeo do caminhar

#### 2.4.1 O ciclo do caminhar

O ciclo do caminhar, como sugere a Figura 8, é dado pelo intervalo de tempo (ou número de quadros/silhuetas) entre solicitações sucessivas do “toque do calcanhar”, que se inicia no instante em que o calcanhar de um dos pés toca o solo e termina quando o calcanhar do mesmo pé faz contato com o solo novamente (STEVENAGE; NIXON; VINCE, 1999); (NOVACHECK, 1998).

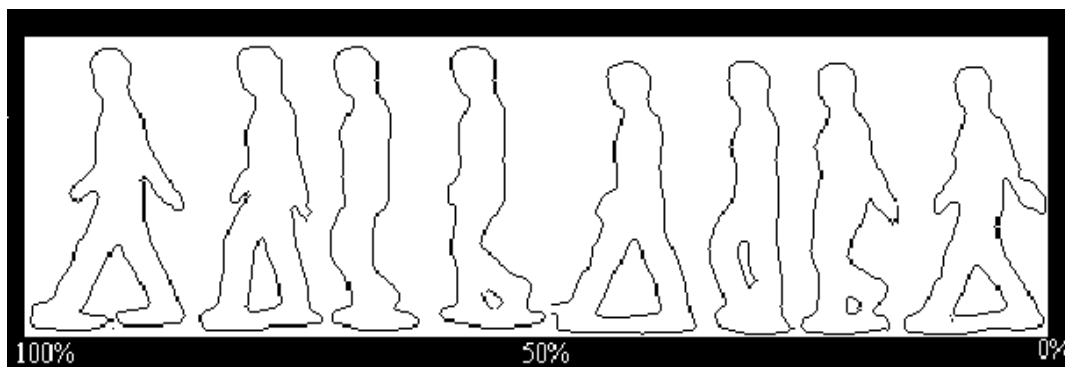


Figura 8 - Seqüência de silhuetas representando um “ciclo do caminhar”

Sendo assim, a forma de caminhar do indivíduo, por apresentar-se de forma cíclica em que a média e variância podem ser consideradas constantes, permite que o tratamento das variáveis envolvidas no modelo possa ser feito através de um

processo estocástico estacionário que leve em conta tão somente um ciclo da mesma.

#### 2.4.2 Métodos para obtenção das características do ciclo do caminhar

As características são obtidas diretamente das seqüências de silhuetas, e, dependendo da abordagem utilizada, podemos ter diversas formas de representação. Assim, em trabalhos cujos autores utilizam abordagens baseadas em modelo, os vetores de características são obtidos de acordo com a representação gráfica dos parâmetros que se apresentam na Figura 9 (BOULGOURIS; HATZINAKOS; PLATANIOTIS, 2005).

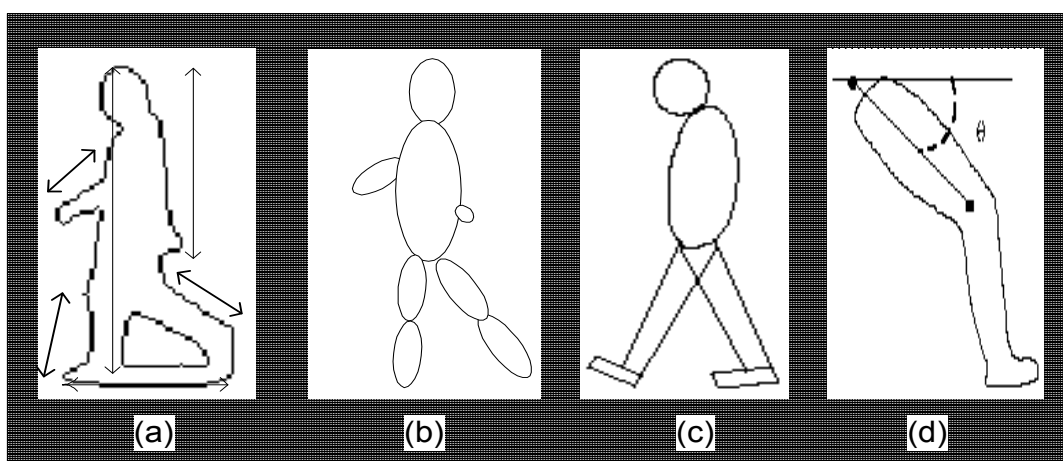


Figura 9 – Extração de características com uso de abordagens baseadas em modelo

Deste modo, temos em: (a) os comprimentos do corpo e dos membros são usados como parâmetros estáticos do caminhar; em (b) as elipses são ajustadas em determinadas partes da silhueta; em (c) tem-se a combinação de diversas formas geométricas sendo ajustadas a determinadas partes da silhueta e, em (d), o ângulo de rotação entre o quadril e a coxa é utilizado.

Por outro lado, em trabalhos cujos autores utilizam abordagens livres de modelo ou holísticas, os vetores de características são obtidos de silhuetas de acordo com o que se apresenta na Figura 10 (BOULGOURIS; HATZINAKOS; PLATANIOTIS, 2005).

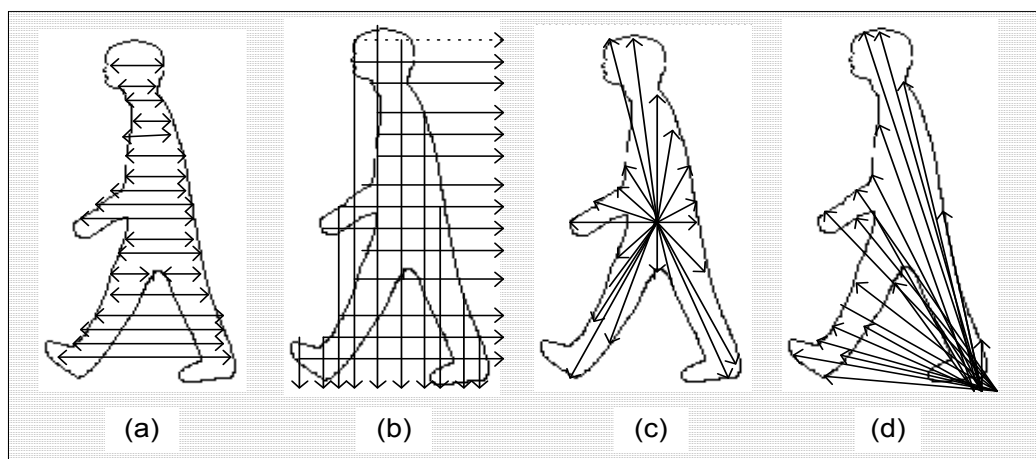


Figura 10 – Extração de características de silhuetas binárias

Neste caso, temos na Figura 10, em (a) a largura da silhueta; (b), as projeções verticais e horizontais dos pontos de contorno da silhueta; (c), as distâncias do contorno ao centro de gravidade da silhueta e, em (d), as distâncias do canto da janela padrão a todos os pontos de contorno da silhueta. Além dessas, há ainda a possibilidade de se utilizarem as silhuetas cheias do ciclo para comporem uma única imagem contendo, em seus pixels, os índices de energia do movimento.

O uso das características obtidas diretamente das silhuetas binárias extraídas das seqüências tem se mostrado inadequado para a tomada de decisões. Por um lado, isso acontece diante da grande quantidade de amostras no vetor de características e, por outro, em função da pequena taxa de discriminação entre as classes. Sendo assim, objetivando selecionar melhor as características dos diversos vetores das seqüências, bem como para permitir uma posterior redução de sua dimensionalidade, técnicas específicas de seleção de características devem ser aplicadas.

# Capítulo 3

## Métodos de Extração, Seleção e Classificação de Características

A fase de extração de características consiste em encontrar uma forma eficiente de representar os dados originais oriundos dos sensores que, em geral, envolve o uso de uma função que seja capaz de mapear o espaço original em um outro espaço de menor dimensão, constituído dos vetores de características. Neste caso, como ocorre em todo sistema de reconhecimento de padrões, após a obtenção das características brutas, é necessário que essas características, em várias etapas, sofram uma série de pré-processamentos, para adequá-los às necessidades das etapas de extração e seleção das características como fase de preparação para a etapa de classificação que é responsável pelo reconhecimento do indivíduo. Sendo assim, as etapas principais são constituídas da extração das características, normalização das informações, seleção das características que mais discriminam as diversas classes e classificação dos vetores de características em função das suas classes (DUDA et al., 2000).

Neste capítulo, por considerar que as técnicas a serem apresentadas, em sua maioria, são amplamente discutidas na literatura, os tópicos relacionados serão expostos de forma resumida, com o único objetivo de permitir uma melhor exposição e facilitar o encadeamento das implementações dos próximos capítulos.

### 3.1 Normalização das características biométricas

A etapa de normalização deve ocorrer logo após a obtenção dos dados brutos, para permitir uma melhor representatividade das características. Três métodos de



normalização mais empregados serão apresentados nesta seção (INDOVINA et al., 2003):

### 3.1.1 Normalização de mínimo-máximo

Esse método mapeia as características obtidas dos sensores no intervalo  $[0, 1]$ , ou seja,

$$n = \frac{s - \min(S)}{\max(S) - \min(S)} \quad (8)$$

onde  $s$  é a característica a ser normalizada, e  $\max(S)$  e  $\min(S)$  especificam os valores extremos das características fornecidas pelos sensores.

### 3.1.2 Normalização de média zero

Esse método transforma as características em uma distribuição com média zero e variância unitária, ou seja,

$$n = \frac{s - m(S)}{\sigma(S)} \quad (9)$$

onde  $S$  é o vetor das características,  $s$  é a característica a ser normalizada,  $m(S)$  é a média e  $\sigma(S)$  é o desvio padrão das características fornecidas pelos sensores.

### 3.1.3 Normalização através da tangente hiperbólica

Esse método está entre aqueles classificados como estatisticamente robusto e foi o método escolhido em todas as nossas implementações. Ele mapeia os valores de cada característica dos sensores em um intervalo  $(0,1)$ , através do cálculo de

$$n = \frac{1}{2} \left[ \tanh \left( 0,01 \frac{s - m(S)}{\sigma(S)} \right) + 1 \right] \quad (10)$$

onde  $S$  é o vetor de características,  $s$  é o atributo a ser normalizado,  $m(S)$  é a média e  $\sigma(S)$  é o desvio padrão das características fornecidos pelos sensores.

Após a etapa de normalização das características obtidos dos sensores, a próxima etapa consiste em aplicar uma entre as diversas técnicas de seleção, mais adequada aos dados a serem classificados. As próximas seções, de forma generalizada, se encarregam de apresentar estas técnicas.

### 3.2 A Técnica da proporção de variâncias

A técnica de seleção de características através da Proporção de Variâncias (*PoV - Proportion of Variances*) permite implementar um procedimento para redução de dimensionalidade. Ela consiste em calcular a razão dada pelas variâncias das amostras inter-classes em relação às variâncias das amostras intra-classes. Portanto, para se aplicar a técnica, considera-se o conjunto dos vetores de características representados pelas linhas de uma matriz  $X$  de dimensões  $(KQ \times M)$ , onde  $K$  é o número de classes;  $Q$  é o número de vetores de características por classe e  $M$  é o número de amostras de cada vetor dispostas em colunas de  $X$ .

Assim, calcula-se os coeficientes dados pela razão entre as referidas variâncias que irão formar o vetor de proporções de variâncias  $r_m$ , em seguida, a partir de um certo valor limiar de  $r_m$ , seleciona-se as colunas de  $X$  cujas amostras apresentam os coeficientes de maior magnitude, que, por sua vez, representam os maiores níveis de discriminação entre as classes.

Em termos matemáticos, ao considerar-se o  $q$ -ésimo ( $q = 1, 2, \dots, Q$ ) vetor de características da  $k$ -ésima ( $k = 1, 2, \dots, K$ ) classe, tem-se que a  $m$ -ésima ( $m = 1, 2, \dots, M$ ) característica é representada por  $x_m(k, q)$ . Neste caso, a média dos vetores de características de cada classe será expressa por:

$$\bar{x}_m(k) = \frac{1}{Q} \sum_q x_m(k, q) \quad . \quad (11)$$

Por outro lado, a média dos vetores de características de toda a base de dados será:

$$\bar{x}_m = \frac{1}{KQ} \sum_k \sum_q x_m(k, q) \quad . \quad (12)$$

Portanto, a razão  $r_m$  entre as variâncias inter-classes e intra-classes para a  $m$ -ésima coluna da matriz  $X$ , sobre os vetores de características representados nas linhas da mesma, será calculada como:

$$r_m = \frac{\sum_k \left( \bar{x}_m(k) - \bar{\bar{x}}_m \right)^2}{\sum_k \sum_q \left( x_m(k, q) - \bar{x}_m(k) \right)^2} \quad (13)$$

Neste caso, o numerador de (13) expressa a variância entre a média da classe  $k$  e a média global das classes; já o denominador representa a soma das variâncias das características dentro de cada classe (EKENEL; SANKUR, 2004).

Em nossas implementações, a serem apresentadas nos próximos capítulos, observamos uma vantagem adicional do uso da técnica PoV ao possibilitar o incremento de novos vetores de características ao conjunto de treinamento. Isto significa que novas classes podem ser adicionadas à base de dados, sem a necessidade de se recalcularem as representações de todas as demais.

### 3.3 A Análise de componentes principais (PCA–Principal Component Analysis)

A Análise de Componentes Principais é uma técnica de seleção de características que permite também efetuar uma redução de dimensionalidade. Ela projeta linearmente características multidimensionais em um espaço de menor dimensionalidade. O método PCA consiste de diversas etapas que permitem obter os auto-vetores da matriz de características.

Dado uma matriz  $C$  em cujas colunas estão contidos os vetores de características, o auto-espaço a ser gerado pelos auto-vetores  $\mu$  e auto-valores  $\lambda$  da matriz estão relacionados por

$$C\mu = \lambda\mu. \quad (14)$$

Os auto-vetores são normalizados e ortogonais de forma que satisfaçam também a condição

$$\mu_i^T \mu_j = \begin{cases} 1 & \text{se } i = j \\ 0 & \text{se } i \neq j \end{cases} \quad (15)$$

Seja  $\Gamma_k$  um vetor de características ordenado lexicograficamente e representando uma classe  $k$  qualquer do conjunto de vetores. Sendo assim, podemos definir o vetor coluna médio por

$$\phi_k = \Gamma_k - \psi \quad (16)$$

onde o vetor coluna médio  $\psi$  será obtido por

$$\psi = \frac{1}{M} \sum_{k=1}^M \Gamma_k. \quad (17)$$

A matriz de covariância pode então ser calculada por

$$C = \frac{1}{M-1} \sum_{k=1}^M \phi_k \phi_k^T \quad (18)$$

onde  $M$  é o número total do conjunto que caracteriza o sub-espço de vetores características, sendo assim, de (14) e considerando que  $\mu_i^T \mu_i = 1$ , após algumas manipulações algébricas temos que

$$\lambda_i = \sum_{k=1}^M (\mu_i^T \phi_k)^2 = \frac{1}{M} \sum_{k=1}^M \text{var}(\mu_i^T \Gamma_k). \quad (19)$$

Portanto, o auto-valor  $i$  representa a variância do conjunto de vetores de características ao longo do eixo descrito pelo auto-vetor correspondente. Neste caso, para atingir uma considerável redução de dimensionalidade e um bom desempenho na classificação, é suficiente considerar apenas um pequeno número de auto-vetores correspondentes aos maiores auto-valores, uma vez que eles expressam uma grande variância no espaço de características do indicador biométrico considerado.

### 3.4 A análise de componentes independentes (ICA - *Independent Component Analysis*)

A Análise de Componentes Independentes é um método de seleção de características baseado em estatística de ordem superior, que permite a extração de variáveis estatisticamente independentes, a partir de uma mistura delas.

Assim, supondo que se dispõe de um conjunto de observações de variáveis aleatórias ( $s_1(t), s_2(t), \dots, s_n(t)$ ) com uma pequena dependência estatística entre elas, basicamente, ICA exerce uma separação de sinais a partir da projeção das matrizes de mistura lineares e dos sinais das fontes, ambos desconhecidos. Portanto, considerando  $A$  uma matriz de mistura;  $S$ , a matriz fonte contendo os vetores estatisticamente independentes em suas linhas e  $X$ , a matriz de dados aleatórios, o sistema ICA pode ser expresso como:

$$X = AS \quad (20)$$

Nesse caso, sob a consideração de que as fontes de sinais podem ser consideradas como estatisticamente independentes e não-Gaussianas, a Análise de Componentes Independentes permite estimar  $A$  e  $S$  a partir da única informação disponível, que é aquela expressa na matriz  $X$ . Portanto, sob determinadas condições sujeitas a algumas restrições, é possível encontrar uma matriz de sinais separados  $W$  ao se maximizar alguma medida de independência dos sinais, tal como a sua entropia. Em outras palavras, podemos escrever (HYVÄRINEN; KARHUNEN; OJA, 2001)

$$U = WX \text{ onde: } W = A^{-1} \text{ e } U \cong S. \quad (21)$$

Na maioria das aplicações, apesar dos sinais não serem completamente independentes em termos estatísticos, ainda assim o método ICA pode estimar os coeficientes de contribuição deles, o que permite efetuar a separação dos sinais originais.

No reconhecimento de pessoas pela sua forma de caminhar, muitas informações importantes podem estar contidas nas relações cinemáticas apresentadas pelas estatísticas de alta ordem das posições dos pixels das silhuetas em uma seqüência. A Análise de Componentes Independentes, então, efetua a separação dos

cumulantes de altas ordens da entrada em relação aos de segunda ordem, como aqueles obtidos com a Análise de Componentes Principais. Portanto, ICA fornece uma representação de dados mais poderosa que PCA, uma vez que sua meta é obter uma representação/decomposição independente estatisticamente entre os dados fontes, ao invés de uma representação descorrelacionada como o PCA.

Para encontrar um conjunto de componentes de características independentes, conforme apresenta a Figura 11, os vetores de características  $X$  são considerados como uma combinação linear de características base  $S$ , independentes estatisticamente, onde  $A$  é a matriz de mistura desconhecida (BARTLETT; MOVELLAN; SEJNOWSKI, 2002).

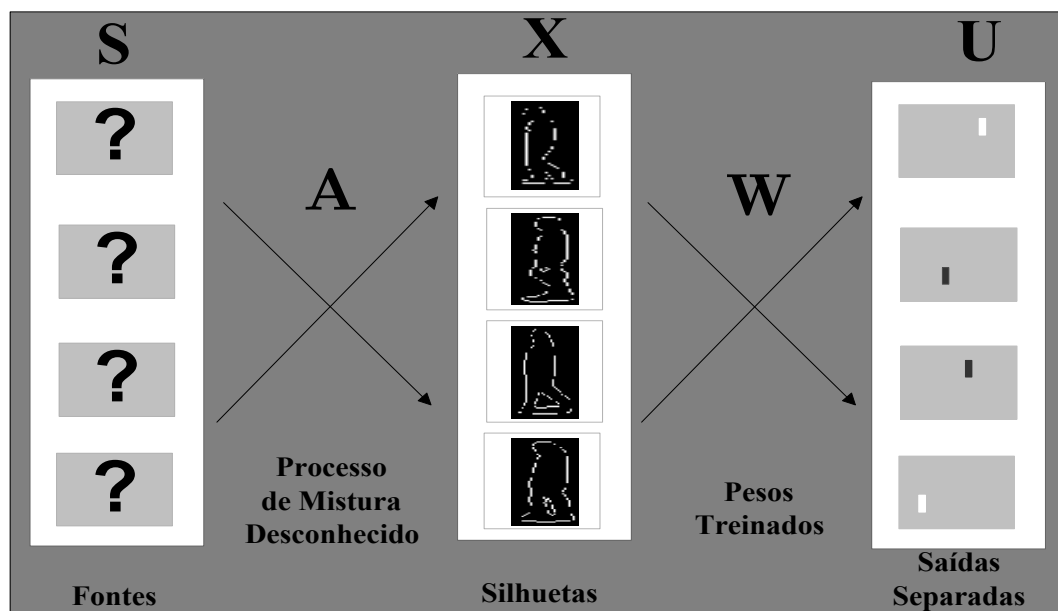


Figura 11 - Modelo ICA simplificado

Assim, as características base, representadas através dos seus vetores de características, são recuperadas a partir do aprendizado dos filtros da matriz  $W$  que, por sua vez, geram as saídas  $U$  estatisticamente independentes.

### 3.5 As transformadas wavelets

Uma wavelet é caracterizada como uma forma de onda com duração efetivamente limitada e que possui média nula. As transformadas wavelets (*WT - Wavelets Transforms*) diferem da análise de um sinal via transformadas de Fourier (*FT - Fourier Transform*), pelo fato das *FTs* representarem, em senóides de mesma

amplitude e frequências distintas, os fenômenos periódicos e invariantes no tempo ou estacionários. Elas também diferem das STFT (*Short Time Fourier Transform*), porque estas utilizam-se de janelas bidimensionais de tempo e frequência, com larguras fixas, para mapear os sinais, permitindo localizar fenômenos de interesse que ocorrem no sinal em determinados intervalos de tempo.

As WTs, por sua vez, são capazes de capturarem características visuais específicas das formas e do interior dos objetos, pois se utilizam de informações extraídas da análise de sinais a partir de janelas ajustáveis em tempo e escala. Portanto, Wavelets, são um tipo de função multi-resolução que efetua uma decomposição hierárquica ou análise de um sinal (FLIEGE, 2000).

A teoria wavelet unidimensional é definida por duas funções ortonormais onde uma delas, a função wavelet mãe, apresenta uma forma que permite acentuar as características relevantes para a análise que se deseja efetuar, e a outra, chamada de função escala, permite ajustar a largura das janelas de observação sobre o sinal ou sobre os vetores de características, proporcionando refinamentos nas frequências em função da posição ocupada pela respectiva janela.

A transformada wavelet é implementada através de filtros espelhados em quadratura, passa-altas (HP) e passa-baixas (LP). Os filtros passa-altas e passa-baixas, quando aplicados aos vetores de características, irão gerar, respectivamente, os coeficientes de detalhes (cD) e de aproximação (cA) do sinal original. Portanto, conforme apresenta a Figura 12 para o algoritmo rápido da transformada wavelet (*FWT – Fast Wavelet Transform*), os níveis de aproximação e detalhes de um sinal são obtidos através deste banco de filtros e de decimações sucessivas aplicadas ao mesmo (BURRUS; GOPINATH; GUO, 1998).

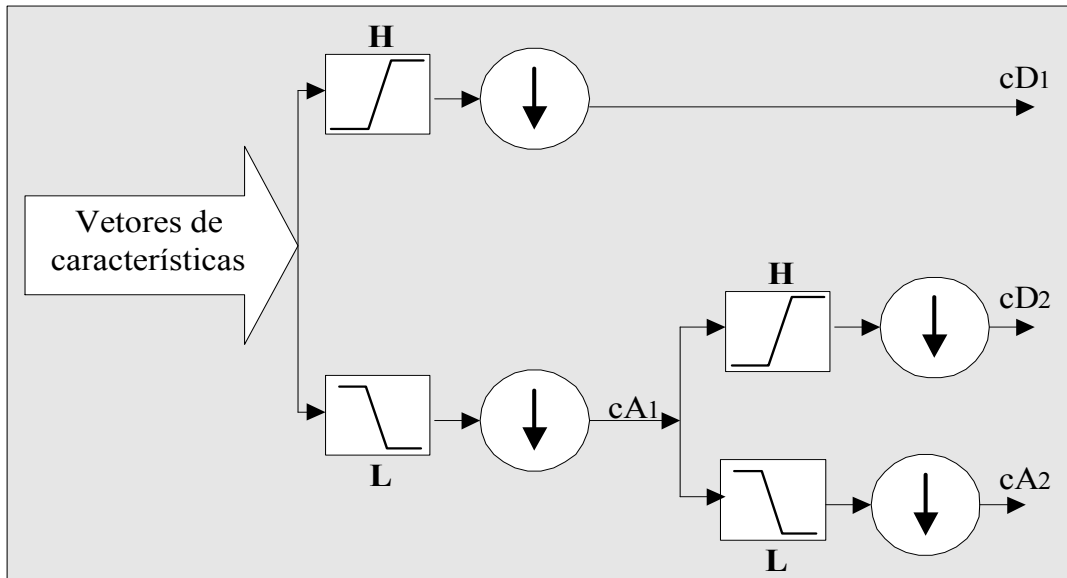


Figura 12 - Decomposição Wavelet dos vetores de características

Assim, características obtidas em diferentes níveis são capazes de gerar coeficientes de aproximação e detalhes diferentes. Para a maioria dos sinais ou características, quando decompostos em diversos níveis, as baixas frequências são as que apresentam as informações mais importantes, isto é, são elas que carregam a maior identidade do sinal, ficando para as altas frequências os ruídos e informações menos importantes.

Em particular, as Wavelets decompõem um sinal dado em uma família de funções com suporte finito. Esta é construída pelas translações e dilatações de uma função chamada Wavelet mãe. O suporte finito da Wavelet mãe localiza o sinal no tempo, enquanto que a função escala permite extrair os componentes de diferentes frequências.

Um espaço importante em processamento de sinais é o chamado espaço de Lebesgue,  $L_2$ , que representa o espaço de todas as funções com convergência quadrática. A análise de multi-resolução é formulada por exigir um “aninhamento” dos espaços estendidos de forma que cada um está contido em outro, tal que seja satisfeita a expressão

$$\dots V_{-1} \subset V_0 \subset V_1 \dots \subset L_2. \quad (22)$$



É possível, então, construir uma análise de multi-resolução e obter níveis de detalhes mais finos no subespaço  $V_{j+1}$  do que em  $V_j$ . Neste caso, as características de um sinal podem ser descritas ou parametrizadas através da definição de um conjunto diferente de funções  $\psi$ , que transpõem a diferença entre os espaços alcançados por várias janelas da função escala. Essas funções são as Wavelets, que transpõem o espaço  $W_j$  tal que  $V_{j+1}=V_j\oplus W_j$  e podem ser descritas como

$$\psi_i^j(x) = 2^{j/2} \psi(2^j x - i). \quad i = 0, \dots, 2^j - 1. \quad (23)$$

Portanto, funções escala e Wavelet diferentes permitirão obter diversas transformadas Wavelets. Sendo assim, há varias espécies de funções que satisfazem as propriedades das Wavelets, sendo que as de Haar e as de Daubechie são consideradas básicas e mais conhecidas. Por ser a mais básica, a primeira família de Wavelets a considerar é a de Haar (SINGH et al., 2004). Sua função escala é dada por

$$\phi(x) = \begin{cases} 1 & \text{se } 0 \leq x < 1 \\ 0 & \text{caso contrário} \end{cases}, \quad (24)$$

e sua wavelet mãe é definida como

$$\psi(x) = \begin{cases} 1 & \text{se } 0 \leq x < 1/2 \\ -1 & \text{se } 1/2 \leq x < 1 \\ 0 & \text{caso contrário} \end{cases}. \quad (25)$$

As Wavelets, quando utilizadas na seleção das características, também possibilitam a implementação da aprendizagem incremental de novas classes. Isto significa que novas características podem ser adicionadas à base de dados, sem a necessidade de se recalcularem as representações de todas as outras já treinadas.

### 3.6 Classificadores baseados em máquinas de vetores suporte

As Máquinas de Vetores Suporte (*SVM – Support Vector Machines*) são máquinas de aprendizagem baseadas na minimização de um risco estrutural obtido a partir do valor esperado do erro das características ou amostras atuais em relação à máquina treinada a partir do conjunto de treinamento.

Seja um conjunto de observações formadas por um par contendo o vetor de características  $\mathbf{x}_i \in R^n$ ,  $i = 1, 2, \dots, l$ , e a sua classe  $y_i$  correspondente. Assumindo que há alguma distribuição de probabilidade desconhecida  $P(\mathbf{x}, y)$  associada com os dados, suponha agora que temos uma máquina cuja tarefa seja aprender a mapear  $\mathbf{x}_i \rightarrow y_i$ , isto é, a máquina é definida por um conjunto de mapeamentos  $\mathbf{x}_i \rightarrow f(\mathbf{x}, \alpha)$ , possíveis, onde tais funções são rotuladas através do parâmetro ajustável  $\alpha$ . Neste caso, o erro esperado entre o vetor teste e a máquina treinada será dado por

$$R(\alpha) = \int \frac{1}{2} |y - f(\mathbf{x}, \alpha)| dP(\mathbf{x}, y). \quad (26)$$

Deste modo,  $R(\alpha)$  é chamado de risco esperado ou risco atual e  $\alpha$  representa todos os parâmetros da máquina treinada. Por outro lado, o “risco empírico” definido como a medida da taxa média do erro em relação ao conjunto de treinamento para um conjunto finito de observações e com o número  $l$  de parâmetros de treinamento, é expresso por

$$R_{emp}(\alpha) = \frac{1}{2l} \sum_{i=1}^l |y_i - f(\mathbf{x}_i, \alpha)|. \quad (27)$$

O risco empírico é um número fixo para uma escolha particular de  $\alpha$  e de um conjunto  $\{\mathbf{x}_i, y_i\}$  de treinamento. O termo  $\frac{1}{2} |y_i - f(\mathbf{x}_i, \alpha)|$  é chamado de perda e para o caso considerado aqui ele pode assumir os valores 0 e 1. Agora considerando algum  $\eta$  tal que  $0 \leq \eta \leq 1$ , então para perdas nesta faixa e para a probabilidade  $1 - \eta$ , o risco esperado será expresso por

$$R(\alpha) \leq R_{emp}(\alpha) + \sqrt{\frac{h(\log(2l/h) + 1 - \log(\eta/4))}{l}}, \quad (28)$$

onde  $R_{emp}$  é o risco empírico, que é dado pelo erro médio sobre o conjunto de treinamento, e  $h$  é chamada dimensão VC (Vapnik-Chervonenkis). Assim, as máquinas de vetores suporte minimizam o segundo termo da Equação 28 para um risco empírico fixado.

Os classificadores de padrões SVM foram basicamente desenvolvidos para problemas de duas classes. Eles superam os classificadores convencionais, especialmente quando o número de dados de treinamento é pequeno e não há sobreposição entre as classes (BURGES, 1998).

### 3.6.1 Classificador SVM para duas classes linearmente separáveis

Para o caso de classes linearmente separáveis, o classificador SVM encontra o hiperplano ótimo (superfície de decisão) que separa os padrões de treinamento, conforme a Figura 13. Esse hiperplano é encontrado de tal forma que maximiza um parâmetro denominado margem. Tal margem é representada pela soma das distâncias, em módulo, dos padrões de treinamento em relação ao hiperplano, para melhor discriminar as classes.

Assim, considerando um conjunto de vetores características  $\mathbf{x}_i$  de dimensão  $M$  pertencente às classes 1 ou 2, cujos rótulos associados são  $\mathbf{y}_i = 1$  para a classe 1 e  $\mathbf{y}_i = -1$  para a classe 2, então, se as classes são linearmente separáveis, pode-se determinar uma função de decisão dada por  $D(\mathbf{x}) = \mathbf{W}^T \mathbf{x} + b$ , onde  $\mathbf{W}$  é um vetor de dimensão  $M$  e  $b$  é um escalar, que irão satisfazer a equação

$$\mathbf{y}_i D(\mathbf{x}_i) \geq 1 \quad \text{para } i = 1, \dots, M. \quad (29)$$

Nesse caso, a distância entre o hiperplano de separação das classes  $D(\mathbf{x}) = 0$  e a amostra do conjunto de treinamento mais próxima do mesmo é chamada de “margem”, e o hiperplano desejado é aquele com a máxima margem, que agora é chamado de hiperplano ótimo de separação. Sendo assim, a distância Euclidiana entre um dado de treinamento  $\mathbf{x}$  e o hiperplano de separação é dada por

$|D(x)| / \|W\|$ . Portanto, se considerarmos uma margem  $\delta$ , todos os dados de treinamento devem satisfazer à condição

$$\frac{y_k D(x_k)}{\|W\|} \geq \delta \quad \text{para} \quad k = 1, \dots, M. \quad (30)$$

Sendo  $W$  uma solução da Equação 30, então  $aW$ , onde  $a$  é um escalar, também o será. Isso permite impor a seguinte restrição:  $\delta \|W\| = 1$ .

Dessa forma, para encontrar o hiperplano ótimo de separação, é necessário encontrar o vetor  $W$  com a mínima norma Euclidiana que satisfaça a igualdade da Equação 30. Neste caso, os pontos de treinamento (amostras) encontrados, que, se removidos, podem alterar a solução encontrada para a margem, são chamados de “vetores suporte”, conforme ilustra a Figura 13(a) para duas classes linearmente separáveis e bidimensionais onde os mesmos são assinalados nos círculos.

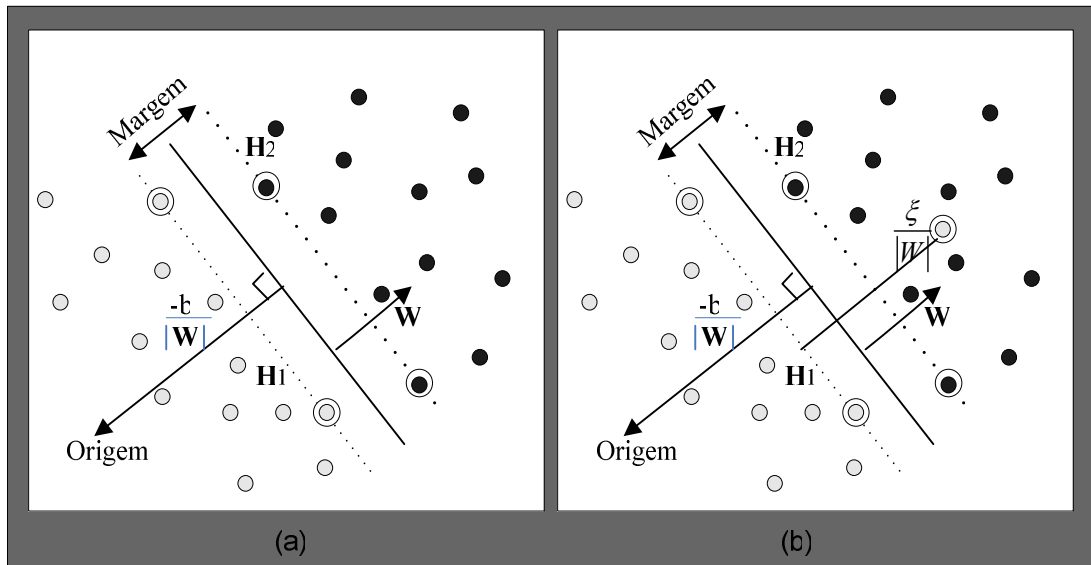


Figura 13 - Hiperplanos de separação para duas classes: a) separáveis b) não-separáveis

A obtenção do hiperplano ótimo de separação é agora possível através da minimização da energia do vetor  $W$ , dada por  $1/2\|W\|^2$ , em relação a  $W$  e a  $b$ , e sujeita à restrição

$$y_i (W^t x_i + b) \geq 1 \quad \text{para} \quad i = 1, \dots, M. \quad (31)$$

### 3.6.2 Classificador SVM para duas classes linearmente não-separáveis

Para o caso de duas classes lineares não-separáveis, ainda assim é possível estabelecer uma solução, ao se introduzir na função risco um parâmetro adicional  $\xi$  capaz de ponderar os erros de classificação. Isso irá provocar uma pequena alteração na Equação 31, modificando-a para

$$y_i (w^t x_i + b) \geq 1 \pm \xi_i \quad \xi_i \geq 0 \quad \text{para } i = 1, \dots, M. \quad (32)$$

A única diferença, agora, em relação ao hiperplano ótimo, é que o mesmo ficará limitado superiormente pelas consequências da alteração na equação de restrição, conforme apresenta a Figura 13(b). Uma generalização na formulação de todos os casos é possível ao se introduzir um multiplicador de Lagrange nas restrições (BURGES, 1998).

Para o caso não-linear, os padrões de treinamento são mapeados em um espaço de alta dimensão através de uma função “kernel”. Isto irá permitir que, neste novo espaço, o limite de decisão seja linear. Assim, o algoritmo de treinamento deve agregar uma “função Kernel” capaz de se adaptar adequadamente às mudanças exigidas. As funções kernel mais utilizadas para aplicações em reconhecimento de padrões são do tipo polinomial de grau  $p$ , gaussiana RBF e tangente hiperbólica, ou seja

$$K(x, y) = (x \cdot y + 1)^p,$$

$$K(x, y) = e^{-\|x \cdot y\|^2 / 2\sigma^2}, \text{ e} \quad (33)$$

$$K(x, y) = \tanh(kx \cdot y - \delta), \text{ respectivamente.}$$

Finalmente, apesar dos classificadores SVM apresentados serem binários, eles podem ser facilmente combinados e adaptados para tratar problemas de múltiplas classes. Uma adaptação simples para atender tal requisito pode ser implementada utilizando-se do treinamento combinado de  $N$  classificadores em uma configuração do tipo “um contra o resto”, isto é, para cada passo do treinamento, uma classe será considerada positiva e todas as demais negativas. Neste caso, um

problema de  $n$ -classes ( $n > 2$ ) é convertido para um outro de  $n$ -duas-classes, de tal forma que para o  $i$ -ésimo problema de duas-classes, a classe  $i$  é separada do restante das demais (ABE; INOUE, 2002).

### 3.7 O classificador baseado em redes neurais

O ser humano nasce com bilhões de neurônios, cada um deles interconectado com outros milhares, formando uma imensa rede de processamento distribuído. A criação e modificação destas conexões garantem ao homem a capacidade de aprender, armazenar, generalizar e utilizar conhecimentos através da experiência por toda a vida.

Com o intuito de imitar a capacidade humana de aprendizagem, através de estruturas semelhantes à rede neural biológica, surgiram as Redes Neurais Artificiais (RNAs). Estas, por sua vez, são expressas em uma forma de programação não algorítmica baseada no processamento paralelo de suas unidades, os neurônios artificiais (HAYKIN, 1994).

A Figura 14 representa um modelo de neurônio artificial. As entradas, representadas pelas variáveis  $a_j$ , correspondem aos dendritos dos neurônios biológicos. As ligações sinápticas, representadas pelos pesos  $w_{j,i}$ , interligam os axônios dos outros neurônios e seus dendritos. A ativação da saída representada por  $a_i$ , que depende de um somatório das entradas ponderadas pelos seus respectivos pesos, dependendo de uma função de ativação utilizada, poderá ocorrer ou não em função de um determinado limiar.

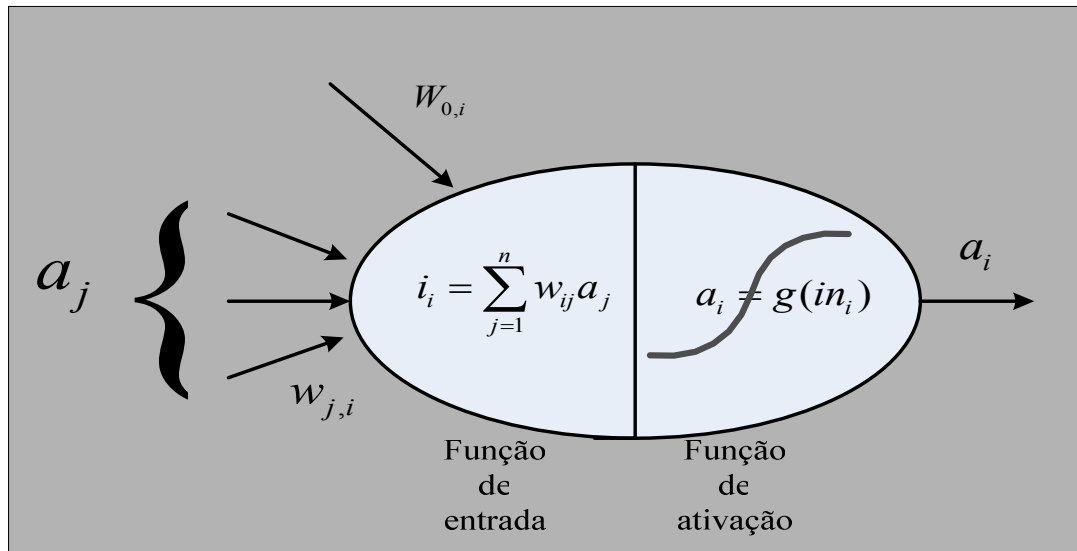


Figura 14 - Modelo de um neurônio artificial

A alteração dos pesos dos neurônios artificiais simula, ainda que de maneira rudimentar, o processo de aprendizado animal, já que a modelagem precisa do cérebro com seus bilhões de neurônios ainda está muito acima da capacidade de qualquer sistema digital existente. Entretanto, ainda assim, a partir do modelo de um neurônio artificial, podem ser construídas as RNAs que são neurônios artificiais interconectados, simulando a estrutura neural biológica, e utilizadas em aplicações específicas.

As redes neurais são bastante utilizadas em aplicações onde um modelo matemático não está disponível. É uma técnica que pode ser facilmente implementada em *hardware*, permitindo a execução de tarefas complexas em tempo real. A capacidade de generalização das redes neurais é uma característica muito interessante para aplicações onde não se têm informações precisas.

### 3.7.1 O classificador neural Bayesiano

Esse classificador utiliza uma abordagem baseada na técnica de “máximos a posteriori” (MAP), cuja função custo é dada pela soma da verossimilhança logarítmica e de um termo de regularização (LARSEN et al., 1998).

Para que o classificador possua uma maior robustez, a função custo é alterada, passando a incorporar o risco de contaminações nos dados, o que leva à introdução de um novo parâmetro de probabilidade. Neste caso, o classificador neural necessita

otimizar os pesos da rede, da probabilidade contaminante e dos parâmetros de regularização.

Para adaptar a probabilidade de contaminação, bem como os parâmetros de regularização, é minimizado o erro sobre um conjunto de validação dos dados, através do gradiente descendente. Além disso, um detector de contaminações é implementado na própria estrutura da rede.

A rede neural multicamadas utilizada no classificador faz uso do aprendizado supervisionado com algoritmo *feed-forward backpropagation*, onde a função de ativação do tipo tangente hiperbólica é usada na camada oculta, e a função softmax é usada na camada de saída (BISHOP, 1995).

Nessa proposta, os pesos são otimizados com a abordagem MAP (Máximo a Posteriori), onde a função de erro do tipo entropia cruzada é associada a uma Gaussiana para anteceder aos mesmos, e a probabilidade de contaminação nos dados é estimada a partir dos exemplos de treinamento (SIGURDSSON et al., 2002).

### **3.7.2 O classificador neural por redes de base radial**

As redes neurais artificiais com funções de base radial têm assumido uma posição significativa dentro do universo das redes neurais artificiais, em função da simplicidade do processo de treinamento e do baixo custo computacional, sendo usadas, principalmente, em problemas de aproximação de funções, predição e classificação. Sua estrutura em múltiplas camadas admite o método de treinamento *feedforward* tanto supervisionado como híbrido. A classe das funções radiais é especial porque sua característica principal é que sua resposta diminui (ou aumenta) monotonicamente com distância de um ponto central.

A Figura 15 apresenta a estrutura de uma rede neural artificial RBF com três camadas. A primeira camada realiza, de forma linear, a conexão do modelo com o meio; a segunda, cujos neurônios artificiais possuem funções de ativação do tipo radiais de base, geralmente do tipo Gaussianas, realiza uma transformação não-linear do espaço vetorial de entrada projetando-o em um outro espaço, que



geralmente possui uma maior dimensão; e a última, transforma de forma linear o espaço vetorial interno na saída do sistema (HAYKIN, 1994).

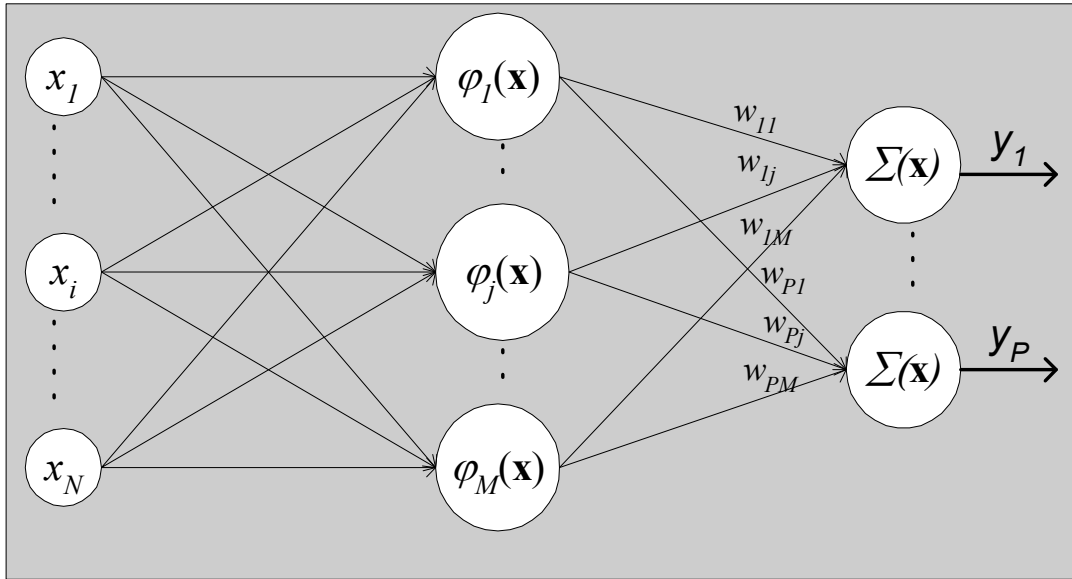


Figura 15 - Estrutura de uma rede RBF

As redes RBFs, baseadas em funções Gaussianas, produzem respostas com magnitudes mais elevadas quando o padrão está dentro de uma pequena região localizada no espaço de entrada. Neste caso, cada função requer genericamente um ponto central  $\mathbf{c}$  e um parâmetro escalar  $v$ , podendo ser expressa por

$$\varphi(v) = e^{\left(-\frac{1}{2\sigma^2}\|v - c\|^2\right)}. \quad (34)$$

Assim, de acordo com a Figura 15, cada componente  $y_p$  do vetor de saída da rede RBF será caracterizado como:

$$y_p = \sum_{j=1}^M w_{pj} \varphi_j(x_i), \quad (35)$$

que, das Equações 34 e 35, após algumas manipulações algébricas, dará como resultado

$$y_p = \sum_{j=1}^M w_{pj} e^{\left(-\frac{1}{2\sigma_j^2}\|x_i - c_j\|^2\right)}, \quad (36)$$

onde  $w_{pj}$  é o peso sináptico entre o neurônio  $j$  da camada escondida com o neurônio  $p$  da saída,  $\mathbf{x}_i$  é o  $i$ -ésimo vetor de entrada do conjunto de treinamento  $\mathbf{X}$ , e  $\mathbf{c}_j$  é o vetor central ligado ao neurônio  $j$  da camada escondida.

# Capítulo 4

## Sistemas de Identificação Pessoal por meio da Forma de Caminhar

A análise biomecânica da forma de caminhar, já há alguns anos, vem tendo uma grande importância nas pesquisas no campo da saúde e, em especial, na Bioengenharia. Assim, é sabido que alterações básicas da marcha, observadas no padrão de caminhar pessoal, podem ser um indicador prévio do início da doença de Parkinson, da esclerose múltipla e da hidrocefalia de pressão normal (MELATO; BIGAL; SPECIAL, 2000).

Recentemente, em consequência das necessidades crescentes exigidas para a segurança pessoal e patrimonial, houve um grande interesse dos pesquisadores no estudo e aplicação de características biométricas e comportamentais no reconhecimento pessoal. Tais pesquisas têm concentrado seus esforços em estudos que permitam a identificação à distância e, de preferência, sem a colaboração da pessoa. A forma humana de caminhar, através das suas características biomecânicas, é responsável por grande número dessas pesquisas (DAVIS; TAYLOR, 2002); (PHILLIPS et al., 2002); (WANG et al., 2003b).

Neste capítulo, para fins de avaliação de desempenho dos classificadores, apresentamos em cada seção as propostas e os resultados obtidos a partir das diversas abordagens aplicadas ao reconhecimento pessoal pela forma de caminhar. A estratégia utilizada para extrair os vetores de características considera inicialmente duas formas de obtenção das silhuetas da sequência de vídeo e duas outras para a etapa de classificação, sendo uma sobre cada silhueta separadamente e, outra, sobre o vetor de características que considera um conjunto de silhuetas de um ciclo do caminhar.

#### 4.1 Extração de dados da seqüência de vídeo

A forma de caminhar pode, estatisticamente, ser considerada "uma atividade cíclica", onde a cada ciclo, constituído de dois passos, são mantidas a média e a variância associada com as silhuetas da seqüência, o que nos permite efetuar o tratamento das variáveis envolvidas como em um modelo de processo estocástico estacionário. Assim, dispondo-se da seqüência de vídeo do caminhante, é necessário extrair-se as diversas características inseridas em um conjunto de silhuetas de um ciclo que são obtidas dos quadros da seqüência.

Nestas condições, serão necessárias várias etapas de pré-processamento de cada seqüência, que se iniciam com a obtenção das imagens de cada quadro e progridem até a obtenção dos vetores de características.

#### 4.2 A detecção da posição de cada silhueta na seqüência de vídeo

A detecção da posição exata das silhuetas em movimento, a partir da seqüência de vídeo, é de suma importância para a obtenção da janela dinâmica de contorno que irá permitir a padronização do referencial dos vetores de características.

O método tradicional da diferença contínua de dois quadros (*CTFD – Continuous Two Frame Difference*), embora não apresentando uma ótima precisão, é utilizado por permitir que a aplicação atenda aos requisitos de tempo real, em função do relativamente baixo custo computacional. Assim, considerando dois quadros consecutivos, conforme apresentam a Figura 16(a) e a Figura 16(b), e aplicando-se um algoritmo com limiar de "binarização", obtém-se a imagem diferença, conforme apresenta a Figura 16(c).

Em seguida, a partir da imagem diferença, faz-se uso de outro algoritmo, que calcula as coordenadas de máxima projeção da imagem sobre os eixos vertical e horizontal. Tais coordenadas irão possibilitar estabelecer uma janela de contenção da silhueta em questão, conforme apresenta o quadro interno da Figura 16(c).

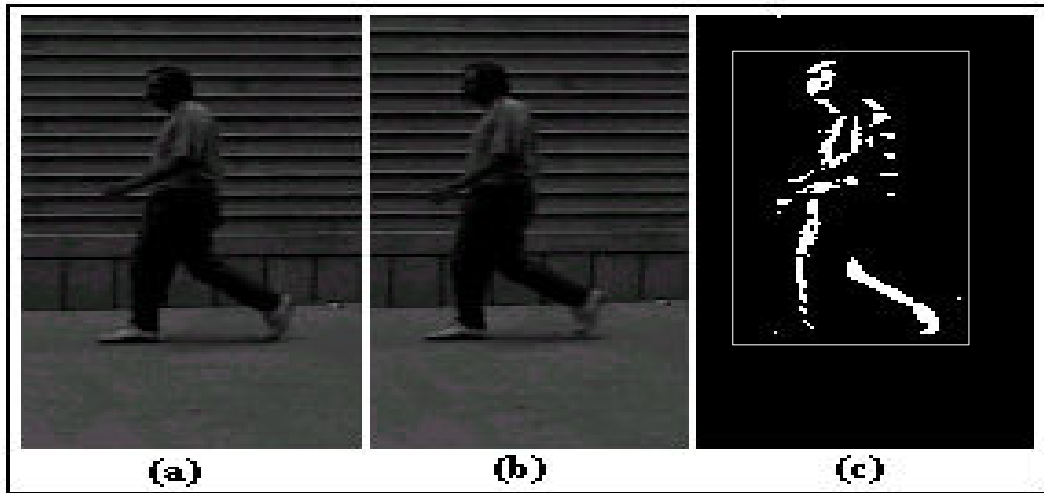


Figura 16 – Obtenção da imagem binária com a janela dinâmica de contorno

### 4.3 A silhueta obtida via subtração do fundo

Considerando que todas as abordagens utilizadas neste trabalho fazem uso das silhuetas interna ou externa, é necessário, para obtê-las, efetuar-se a subtração do fundo de cada uma das imagens de quadros da seqüência.

Considerando-se, também, que o indivíduo, ao caminhar, provoca alterações devidas às regiões dinâmicas de sombras, um método de obtenção de fundo com compensação de brilho baseado na correlação imagem-fundo deve ser utilizado. Para tanto, aplica-se o método do menor erro quadrático médio na escolha de cada “pixel” do conjunto constituído por algumas imagens do fundo e da própria seqüência de vídeo do caminhante. O algoritmo é tal que escolhe, para compor o fundo dinâmico desejado, a posição e a tonalidade do pixel que apresentar o menor erro quadrático médio entre todos os quadros da seqüência. Nesta situação, sendo  $I$  uma imagem de uma seqüência de  $N$  imagens do ciclo, o fundo resultante será dado por

$$f_{xy} = \min_p [med_i (I_{xy}^i - p)^2], \quad i = 1, \dots, N, \quad (37)$$

onde  $p$  é o valor da tonalidade de cinza a ser obtida para o pixel do fundo resultante na posição  $(x,y)$ , e  $i$  representa o índice dos quadros na seqüência.

Na etapa seguinte, para se obter a silhueta de cada quadro em níveis de cinza, efetua-se a subtração pixel-a-pixel de cada frame de uma seqüência em relação ao

fundo. Tal objetivo é atingido com a função de compensação do brilho (WANG et al., 2003b)

$$S(a, b) = 1 - \frac{2 \cdot \sqrt{(a+1)(b+1)}}{(a+1) + (b+1)} \cdot \frac{2 \cdot \sqrt{(256-a)(256-b)}}{(256-a) + (256-b)}, \quad (38)$$

$$0 \leq S(a, b) < 1 \quad e \quad 0 \leq a(x, y), b(x, y) < 255,$$

onde  $a(x, y)$  é a tonalidade do quadro atual e  $b(x, y)$  é a tonalidade do fundo na posição  $(x, y)$ .

Considerando, ainda, a existência de ruídos na imagem diferença, aplica-se a filtragem da mediana para suavizá-la, seguido por um limiar de binarização associado com a conexão dos componentes vizinhos. Em seguida, aplicam-se os operadores morfológicos de dilatação e erosão, com o propósito de eliminar possíveis “buracos” dentro da silhueta, obtendo-se a silhueta cheia, como mostra a Figura 17(a).

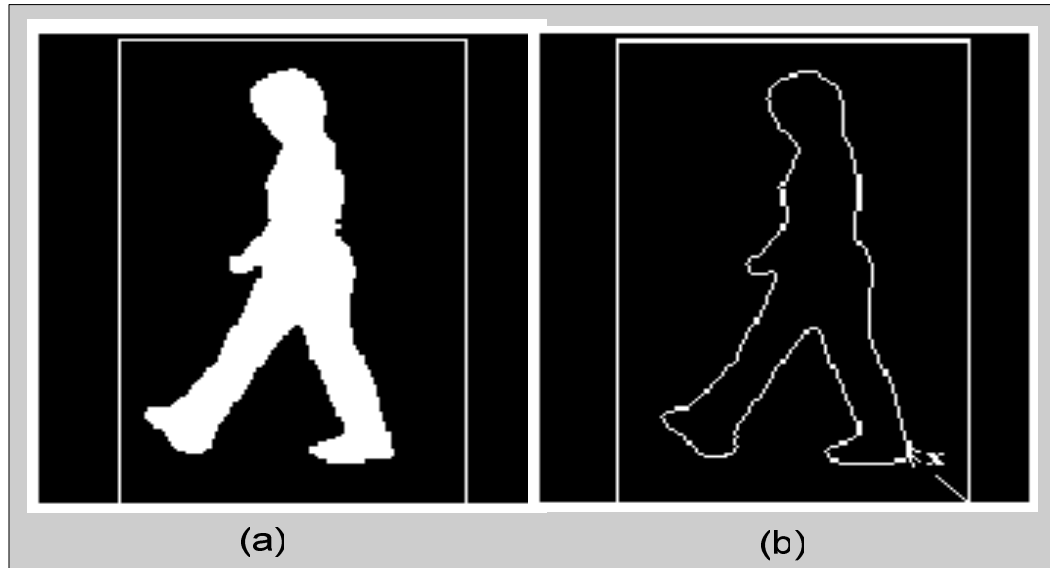


Figura 17 – Janela contendo as silhuetas interna e externa

#### 4.4 A silhueta externa e o vetor de características

A silhueta externa, que irá compor o vetor de características, é extraída através da silhueta cheia, aplicando-se um detector de bordas do tipo Canny (GONZALEZ; WOODS, 2002), conforme apresenta a Figura 17(b). Na etapa seguinte, o contorno

da silhueta e a janela dinâmica, obtidas de acordo com a seção 4.3, permitem extrair o vetor de características  $x$  como as distâncias medidas no sentido anti-horário dos “pixels”.

Nesse ponto, como as silhuetas de um indivíduo ao caminhar sofrem alterações em seus tamanhos, é necessário usar um processo de interpolação-decimação, com passo variável, sobre os vetores obtidos, para padronizá-los.

Nas seções que se seguem, com o objetivo de avaliação de desempenho dos classificadores, apresentamos algumas implementações de reconhecimento pessoal através da forma de caminhar, com abordagens que utilizam métodos e técnicas distintas na extração, seleção e classificação de características.

#### **4.5 Implementação com abordagem Wavelets-PCA e classificador neural**

A nossa primeira implementação, conforme apresenta a Figura 18, inicia-se com a extração dos quadros de diversas seqüências de vídeo do indivíduo caminhando de forma natural. A seguir, calcula-se a imagem média do fundo e, considerando-se a seqüência de imagens, efetua-se a subtração e filtragem entre cada quadro e o fundo para se obter as silhuetas da seqüência.

Em seguida, cada ponto do contorno da silhueta é representado por um vetor de características formado pelas distâncias com origem no canto inferior direito da janela dinâmica que contém a silhueta padronizada.

Neste ponto, é importante destacar que, devido às variações nos tamanhos dos indivíduos, bem como na abertura das pernas de uma pessoa ao caminhar, as silhuetas apresentam diferentes tamanhos. Sendo assim, para padronizá-las foi inserida, no algoritmo de extração do vetor básico de características, uma etapa de decimação que ajusta o passo de interpolação das posições dos pixels a serem suprimidos no contorno das silhuetas.

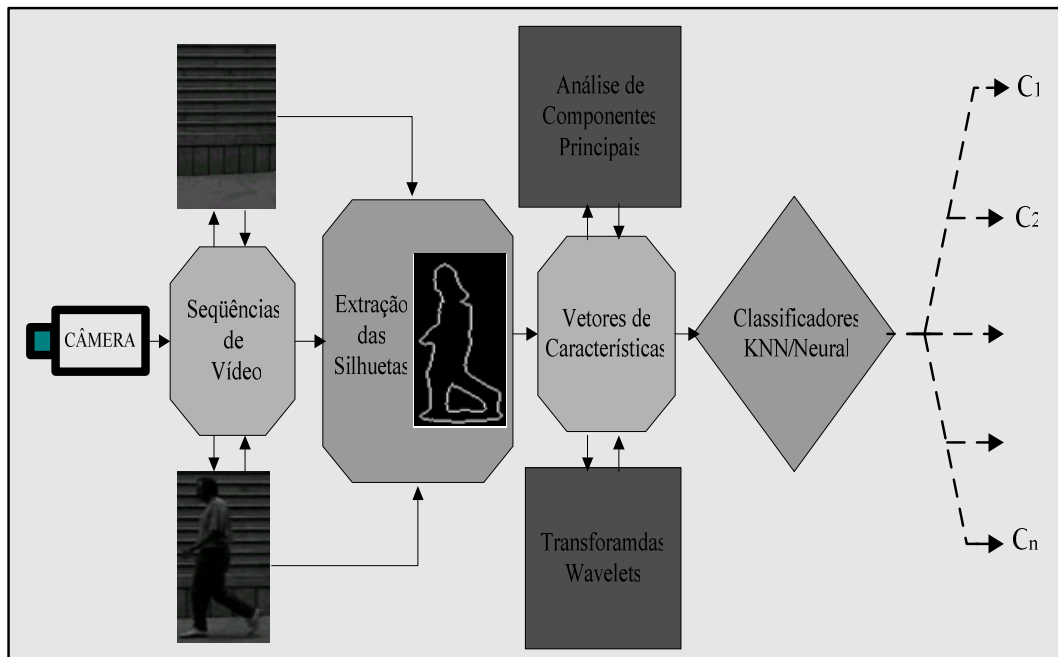


Figura 18 - A abordagem utilizada

A etapa anterior, além de representar uma etapa prévia de redução de dimensão no vetor, permitirá a representação de todas as silhuetas em uma matriz cujas linhas ou colunas serão constituídas dos respectivos vetores de características (COLLINS; GROSS, 2002). Em seguida, em função das informações redundantes e para extrair os coeficientes de maior energia, aos vetores de características aplicam-se a transformada wavelet (WT) e a análise de componentes principais (PCA).

Na etapa final, objetivando-se avaliar e comparar seus desempenhos, utilizam-se de dois classificadores baseados em Rede Neural e nos K-vizinhos mais próximos com similaridade do tipo co-senoidal (*K-Nearest Neighbor- K-NN*).

#### 4.5.1 Os dados e o procedimento experimental

A validação da nossa proposta foi feita com dois bancos de dados de domínio público, obtidos em ambiente natural, com piso de cimento ou de blocos e com uma câmara estática que captura lateralmente as imagens da seqüência, a saber,

- A base de dados UCSD (University of California - San Diego), que foi criada por Little e Boyd (1998), possui 6 indivíduos caminhando em vista lateral com 7 seqüências de vídeo para cada um, obtidas a 30 quadros por segundo. As



42 seqüências de vídeo têm uma média de 65 quadros por seqüência, e cada quadro extraído de uma seqüência tem uma resolução de 320 x 160 pixels, conforme alguns exemplares mostrados na Figura 19(a);

- a base de dados NLPR/CASIA (National Laboratory of Pattern Recognition/Chinese Academy of Sciences - Institute of Automation), que tem 20 sujeitos com 4 seqüências para cada uma das vistas a 0° (perfil), 45° (oblíqua) e 90° (frontal). As 240 seqüências de vídeo foram obtidas a 25 quadros por segundo, e possuem entre 50 a 75 quadros com resolução de 352 x 240 pixels para cada seqüência, conforme alguns exemplares mostrados na Figura 19(b).

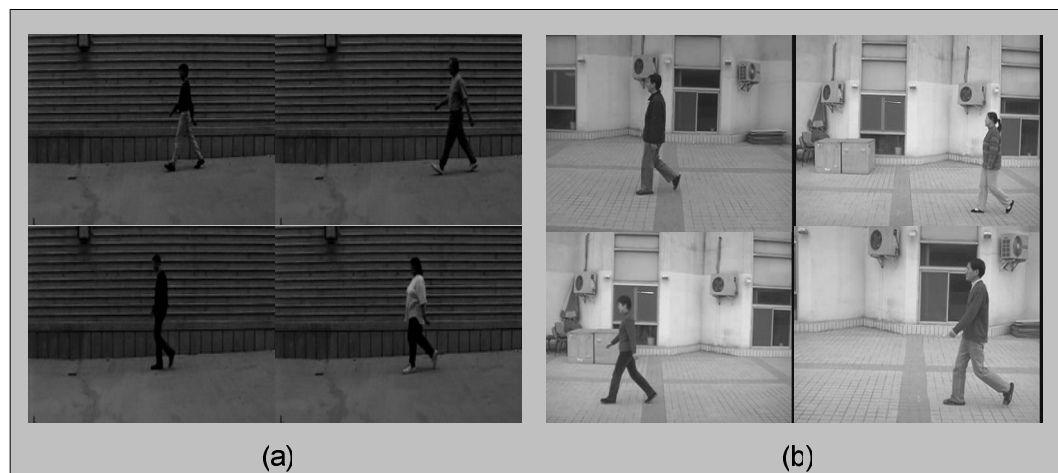


Figura 19 – Exemplos de quadros de seqüências utilizadas das bases de dados

Na implementação da base de dados UCSD, usamos amostras de 6 indivíduos, com 5 seqüências de 33 silhuetas cada por pessoa, totalizando 990 silhuetas. E da base de dados NLPR/CASIA, usamos amostras de 19 sujeitos em vista lateral, com 4 seqüências de 30 silhuetas cada, totalizando 2280 silhuetas. Além disso, utilizamos dois classificadores, o primeiro do tipo “K-vizinhos mais próximos” (DUDA et al., 2000) com similaridade co-senoidal (BARTLETT; MOVELLAN; SEJNOWSKI, 2002) e o outro, um classificador “neural baseado na verossimilhança Bayesiana”, para a obtenção dos resultados.

O classificador neural foi configurado com 12 neurônios na camada escondida e 6 na camada de saída, para atender à base de dados UCSD, e com 38 neurônios na camada escondida e 19 na camada de saída, para a base NLPR/CASIA.

A classificação, para fins de comparação de desempenho dos classificadores, foi realizada considerando cada silhueta, isoladamente, como uma classe correspondente à identidade da pessoa. Ela foi exercida sobre os percentuais de 1, 5 e 10%, respectivamente do total de silhuetas que ficaram de fora da fase de treinamento, isto é, aproximadamente 10, 50 e 100 silhuetas para a base de dados UCSD e 23, 115 e 230 silhuetas para a base NLPR/CASIA. Os resultados de classificação, com uso alternado dos algoritmos K-vizinhos mais próximos (K) e Neural (N), são mostrados na Tabela 2.

Tabela 2 - Resultados dos classificadores sobre o total de vetores

B.D.	CW			PCA/CW			CLAS.
	1%	5%	10%	1%	5%	10%	
UCSD	85	82	88	86	88	92	K
	89	88	87	90	92	89	N
NLPR	83	86	86	85	89	88	K
	87	85	89	90	86	91	N

Neste caso, os resultados são apresentados em conjuntos de três colunas. Assim, aquelas três que se apresentam logo abaixo dos coeficientes Wavelets (CW) contêm os acertos percentuais correspondentes à aplicação da transformada Wavelet Haar - nível 3; e as outras três à direita apresentam os resultados ao ser aplicado o PCA sobre os coeficientes Wavelets (PCA/CW).

Avaliando os resultados, observamos que o desempenho obtido na classificação, quando se efetua de forma conjunta a seleção de características através das técnicas Wavelets/PCA, é ligeiramente superior àquele obtido quando se utiliza a transformada Wavelet.

Uma vez que as silhuetas constituem o parâmetro principal na identidade do indivíduo, e com o objetivo de avaliar o número de silhuetas classificadas corretamente para cada seqüência de vídeo, apresentamos, na Tabela 3, as taxas de acertos, considerando-se 10% do total de silhuetas escolhidas aleatoriamente e utilizadas nos testes para as seis classes da base de dados UCSD.

Tabela 3 - Resultados por classe do classificador neural

Classes	Amostras por classe	Acertos (%)	Acertos (%)
		WT	WT-PCA
1	17	70.0	76.5
2	25	84.0	88.0
3	16	87.5	93.7
4	10	90.0	90.0
5	18	66.6	77.7
6	14	86.7	85.7
<b>Total</b>	<b>100</b>	<b>80.0</b>	<b>85.0</b>

Neste caso, observamos que houve uma taxa média de acertos de, no mínimo, 80% das silhuetas isoladas para as seqüências, o que nos permite, como uma avaliação preliminar, concluir que, provavelmente, o reconhecimento pessoal pela forma de caminhar apresentará um melhor desempenho se forem considerados vetores de características formados através da integração (concatenação) de características de todas as silhuetas de uma seqüência de vídeo.

#### 4.6 Implementação com abordagem wavelets e classificador SVM

Nesta segunda implementação, ainda trabalhamos com as silhuetas isoladamente e utilizamos as transformadas Wavelets na seleção de características. Entretanto, o classificador a ser avaliado é baseado em máquinas de vetores suporte (*SVM – Support Vector Machines*) (ABE; INOUE,2002).

A Figura 20 apresenta os principais blocos da arquitetura a ser utilizada. Neste caso, o processo inicia-se com a captura de quadros a partir das seqüências de vídeo do indivíduo caminhando de forma natural diante de uma câmara.

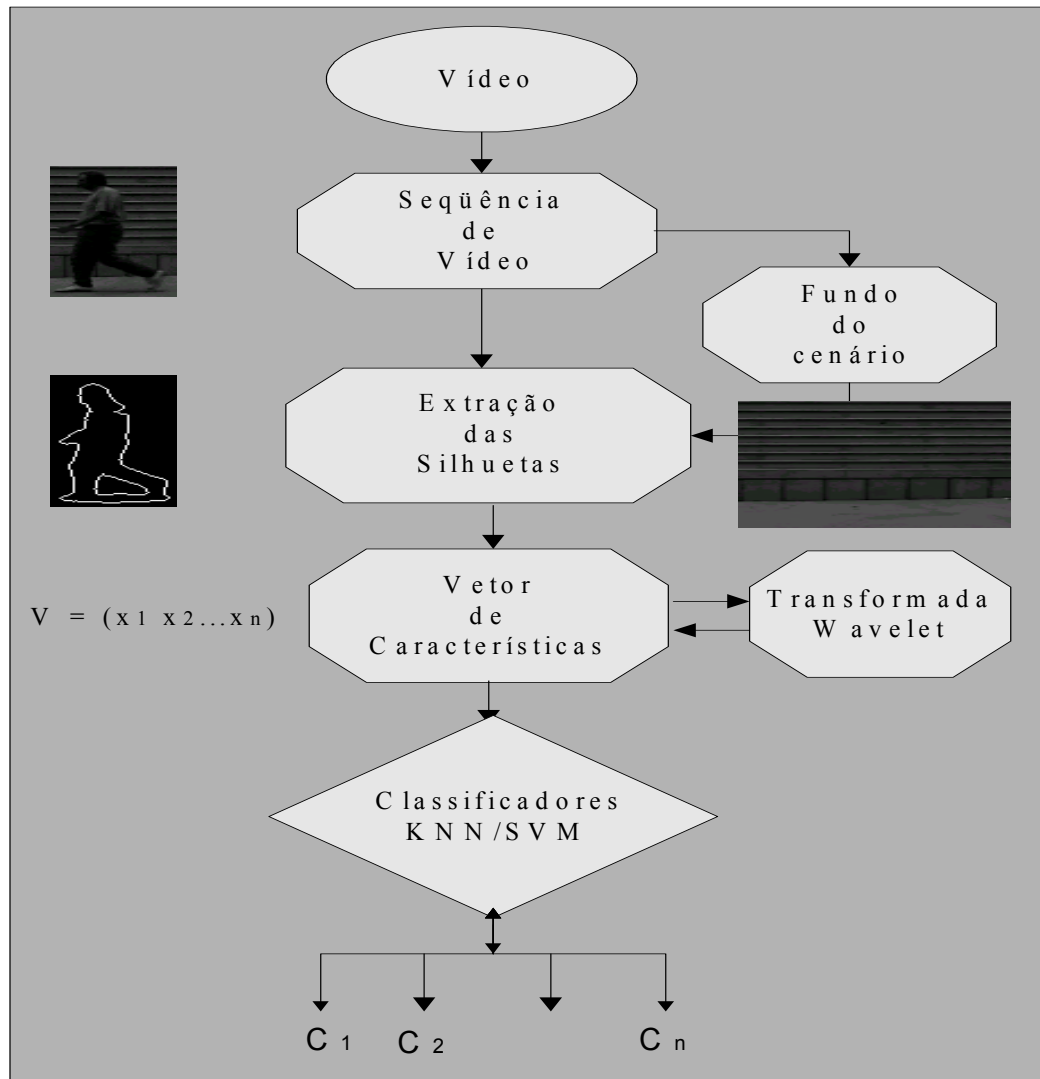


Figura 20 - A arquitetura proposta

Os procedimentos para a obtenção das silhuetas e dos seus respectivos vetores de características são semelhantes àqueles utilizados na implementação da seção anterior. Ao conjunto de vetores de características obtido, pelo fato de possuírem informações redundantes, bem como objetivando reduzir a sua dimensionalidade, aplica-se a transformada Wavelet para extrair os coeficientes de maior energia e representatividade entre as classes (SALOMÃO; ALMEIDA; SALLES, 2005a).

Na etapa final, utiliza-se um classificador SVM e efetua-se a comparação dos resultados obtidos com aqueles de outro classificador K-vizinhos mais próximos, para avaliação de desempenho do primeiro quando utilizado neste tipo de problema.

#### 4.6.1 A base de dados e o procedimento experimental

O banco de dados de domínio público UCSD (LITTLE; BOYD, 1998), obtido em ambiente natural com piso de cimento, foi novamente utilizado aqui para a validação da nossa proposta. Ele possui 6 indivíduos que caminham em vista lateral diante de uma câmara estática. São obtidas 7 seqüências com média de 65 silhuetas para cada, capturadas à razão de 30 quadros/s, totalizando, assim, 42 seqüências.

Na implementação, 5 seqüências de 33 silhuetas cada por pessoa, totalizando 990 silhuetas, foram utilizadas. Utilizamos, também, o classificador “K-vizinhos mais próximos”, por ser de uso clássico em reconhecimento de padrões, para permitir a comparação do seu desempenho com o desempenho do classificador “SVM”.

Em ambos os classificadores, conforme apresenta a Figura 21, utilizamos uma configuração de validação cruzada para permitir a obtenção dos “escores” de identificação das seqüências (CAWLEY; TALBOT, 2003). Neste caso, o conjunto de treinamento é formado por todos os vetores-silhuetas das seqüências, exceto um de teste, que fica de fora, e que se alterna desde a primeira até a última, atingindo, assim, o total delas para cada ciclo do treinamento-validação.

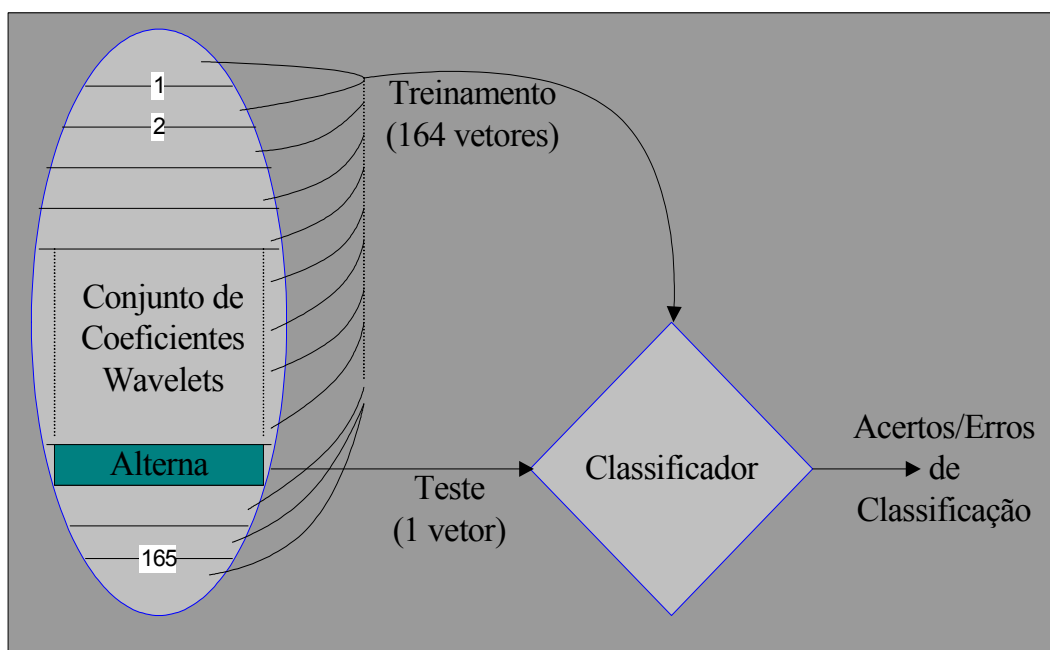


Figura 21 - Processo de treinamento em configuração de “validação cruzada”

O algoritmo foi implementado de tal forma que, antes de “embaralhar” o conjunto de treinamento para cada ciclo, a silhueta que fica de fora para fins de teste corresponde à primeira da classe 1 na seqüência 1 e evolui até a última (165) para a seqüência 5 da mesma classe. Esse procedimento continua para todas as silhuetas com suas classes e seqüências, de tal forma que as silhuetas são testadas na mesma ordem em que aparecem nas seqüências.

Os resultados são apresentados na forma de matrizes de confusão, em que cada linha representa a classe verdadeira ou a seqüência apresentada, e cada coluna representa a classe rotulada pelos classificadores.

A Tabela 4 apresenta, na forma de uma matriz de confusão, os resultados obtidos para 165 silhuetas das 5 seqüências de cada uma das 6 classes, considerando-se o classificador K-vizinho mais próximo e o protocolo de validação-cruzada.

Tabela 4 - Matriz de confusão: classificador K-NN

CLASSES		SAÍDA DO K-NN						PRECISÃO (%)
		1	2	3	4	5	6	
R E A L	1	138	1	6	8	7	5	83,63
	2	0	153	2	4	3	3	92,72
	3	3	1	154	2	3	2	93,33
	4	2	3	5	151	3	2	91,52
	5	1	7	5	1	146	6	87,85
	6	2	1	6	2	6	147	89,09
CONFIANÇA		94,5	92,2	86,5	89,9	86,9	89,1	Total: 89,80

A Tabela 5, por sua vez, apresenta uma outra matriz de confusão com os resultados obtidos para a mesma base de dados considerando-se, agora, o classificador SVM e o protocolo de validação-cruzada para a mesma situação descrita no parágrafo anterior.

Tabela 5 - Matriz de confusão: classificador SVM

CLASSES		SAÍDA DO SVM						PRECISÃO (%)
		1	2	3	4	5	6	
R	1	143	4	5	6	5	2	86,67
	2	3	159	2	1	0	0	96,36
E	3	2	2	158	2	0	1	95,76
A	4	1	1	3	158	0	2	95,76
L	5	5	5	3	1	150	2	90,90
	6	1	1	8	1	2	151	91,52
CONFIANÇA		92,3	92,4	88,3	93,5	95,5	95,6	Total: 92,83

A partir dos resultados apresentados pelos dois classificadores, observa-se um elevado percentual de silhuetas classificadas corretamente por seqüência, o que mostra a viabilidade da proposta na identificação pessoal. Além disso, ao se comparar os resultados das duas tabelas, observa-se que o desempenho do classificador SVM, para todas as silhuetas, foi ligeiramente superior (cerca de 3,4%) ao do classificador K-NN com similaridade co-senoidal.

Para permitir uma avaliação mais detalhada dos acertos de silhuetas por seqüência, a Tabela 6 apresenta, considerando uma situação similar àquela descrita nos parágrafos anteriores, os resultados obtidos considerando-se somente o resultado da classificação “SVM” para as cinco seqüências de cada classe.

Neste caso, com o objetivo de se aproximar o máximo possível de uma situação real, em que a pessoa caminha de forma natural em um cenário diante de uma câmara estática, cada silhueta teste é apresentada ao classificador na ordem em que aparece na seqüência, desde a primeira até a última.

Tabela 6 - Desempenho SVM por seqüência

CLASSES	SEQ.	SAÍDA DO CLASSIFICADOR SVM					
		1	2	3	4	5	6
1	1	27	0	2	0	2	2
	2	32	1	0	0	0	0
	3	29	0	2	2	0	0
	4	28	1	0	2	2	0
	5	27	2	1	2	1	0
2	1	0	33	0	0	0	0
	2	2	31	0	0	0	0
	3	1	30	1	1	0	0
	4	0	32	1	0	0	0
	5	0	33	0	0	0	0
3	1	0	1	30	1	0	1
	2	0	0	33	0	0	0
	3	0	0	33	0	0	0
	4	1	0	32	0	0	0
	5	1	1	30	1	0	0
4	1	0	0	0	32	0	1
	2	0	0	1	32	0	0
	3	0	0	0	32	0	1
	4	0	0	1	32	0	0
	5	1	1	1	30	0	0
5	1	0	0	2	0	31	0
	2	0	1	0	0	32	0
	3	1	0	0	0	32	0
	4	1	1	0	1	30	0
	5	3	3	1	0	25	1
6	1	0	0	1	0	0	32
	2	0	0	2	0	0	31
	3	0	0	1	1	0	31
	4	0	0	3	0	2	28
	5	1	1	1	0	0	30



Ao se analisar os resultados da Tabela 6, pode-se observar que o menor percentual de acertos por seqüência de silhuetas (75,75%), que ocorreu na 5ª seqüência da classe 5, supera em muito a maioria absoluta de silhuetas reconhecidas por seqüência, o que nos dá uma certa garantia de que as cinco seqüências, contendo 33 silhuetas cada e obtidas para as 6 pessoas da base de dados UCSD, foram reconhecidas corretamente.

#### 4.7 Implementação com abordagem PoV-ICA e classificador SVM

Nessa implementação passamos a considerar, no desenvolvimento do nosso sistema automático de reconhecimento pessoal, a utilização das seqüências do caminhar contendo todas as silhuetas de um ciclo na composição dos vetores de características. Isto porque os movimentos do corpo humano, ao caminhar, são melhor caracterizados pelos parâmetros estatísticos gerados a partir dos padrões espaço-temporais extraídos das seqüências de imagens das silhuetas da pessoa.

O sistema proposto, conforme apresenta o diagrama de blocos da Figura 22, inicia-se com a captura de quadros a partir das seqüências de vídeo do indivíduo que caminha lateralmente e de forma natural diante de uma câmara estática.

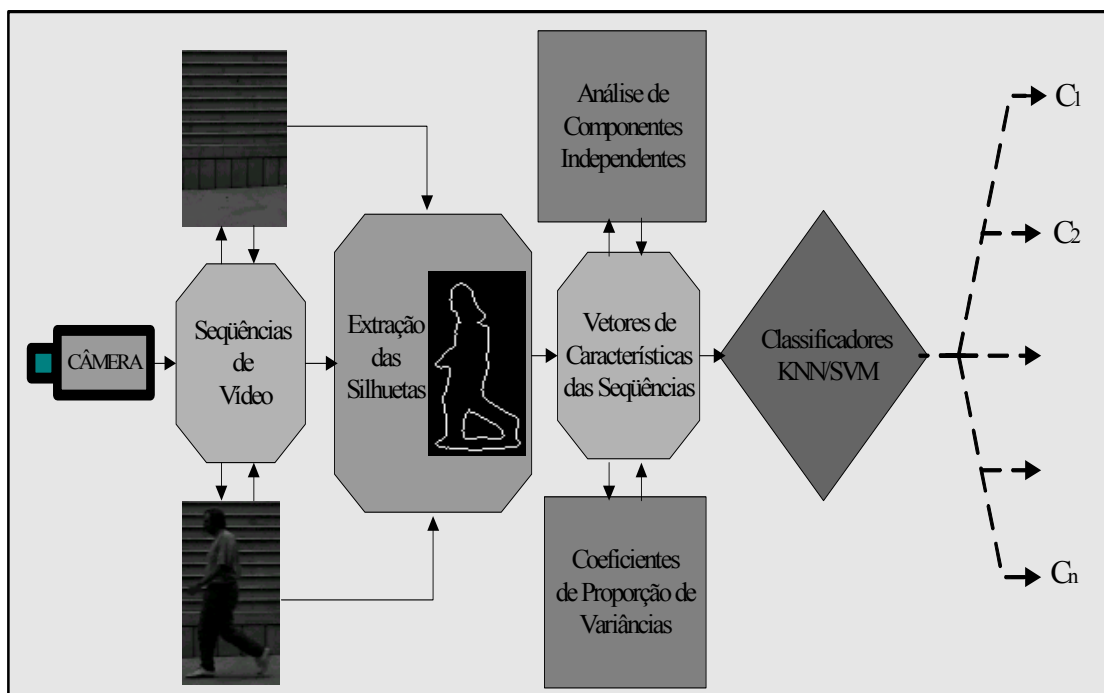


Figura 22 – Síntese em diagramas de blocos da proposta utilizada

Os quadros capturados irão permitir a extração das silhuetas que, por sua vez, irão dar origem às imagens binárias das silhuetas externas, obtidas através de algoritmos detectores/seguidores de bordas. Tal procedimento permite diminuir a dimensão de cada vetor de características a ser obtido e, por conseguinte, o custo computacional envolvido nas etapas seguintes. Em seguida, cada silhueta da seqüência tem suas características extraídas e representadas por um vetor de características obtido pela distância do contorno da silhueta padronizada ao ponto de origem da janela dinâmica que envolve cada uma delas (SALOMÃO; ALMEIDA; SALLES, 2005b).

Ao conjunto de vetores de características, objetivando uma melhor seleção das mesmas, aplicam-se, de forma combinada, a técnica da Proporção de Variâncias (PoV - Proportion of Variance) e o método da Análise dos Componentes Independentes (ICA - Independent Component Analysis).

Na etapa final, utilizam-se dois classificadores, um com a arquitetura baseada em Máquinas de Vetores Suporte e outro, clássico em aplicações de reconhecimento de padrões, do tipo K-vizinhos mais próximos, usado para permitir a avaliação de desempenho do primeiro.

#### **4.7.1 A extração e seleção de características**

Em nossa implementação, utilizamos as 6 pessoas da base de dados UCSD, contendo 7 seqüências de vídeo cada, e padronizamos em 200 pontos por silhueta a dimensão de cada vetor de características. Em seguida, foram concatenadas, na ordem em que aparecem, as 33 silhuetas de um ciclo da seqüência em um único vetor de características, que ficou então com uma dimensão de 6600 pontos. Dessa forma, o conjunto original de vetores de características das silhuetas ficou expresso pelas linhas de uma matriz  $X$  de características com dimensão  $42 \times 6600$ .

A técnica de Proporção de Variâncias foi aplicada ao conjunto de vetores de características expressos nas linhas da matriz, obtendo o vetor de coeficientes  $r_m$ , que expressa a proporção das variâncias inter e intra-classes. Em função desses coeficientes, e considerando um limiar de corte definido em função de valores percentuais de  $r_m$ , para uma primeira etapa de redução de dimensionalidade, foram selecionadas as colunas de amostras da matriz  $X$  mais discriminatórias, eliminando,

do conjunto de dados, as piores amostras relacionadas com os coeficientes menos discriminatórios, conforme apresenta a Tabela 7.

Este procedimento, além de provocar uma primeira redução de dimensionalidade, contribuiu de forma significativa para melhorar a seleção de coeficientes dos componentes independentes, bem como com o desempenho dos classificadores.

Tabela 7 - Limiar percentual do vetor  $r_m$  e redução do número de amostras

<b>Técnica PoV aplicada na pré-seleção de características</b>											
<b>Limiar de <math>r_m</math> (%)</b>	0	1,0	2,0	3,0	4,0	5,0	6,0	7,0	8,0	9,0	10
<b>Nº de Amostras</b>	6600	6075	4717	3350	2242	1502	1051	752	538	387	295

A outra etapa de seleção de características foi feita utilizando-se o método da Análise dos Componentes Independentes, e, para tanto, foi escolhido o algoritmo INFOMAX (BELL; SEJNOWSKI, 1995). Ele é derivado do princípio ótimo da transferência de informação entre neurônios com funções de transferências do tipo sigmoidais, onde, tipicamente, são utilizadas as funções logísticas da forma

$$f_i(u) = \frac{1}{1 + e^{-u}}. \quad (39)$$

A meta do algoritmo é maximizar a informação mútua entre cada vetor de características expresso nas linhas da matriz  $X$  e a saída da rede neural. Esse objetivo é atingido ao se aplicar o gradiente ascendente sobre a entropia da saída em relação à matriz de pesos. Assim, quando há entradas e saídas múltiplas, o procedimento que maximiza a entropia conjunta da saída faz com que cada saída individual movimente-se para uma situação de maior independência estatística.

A implementação faz uso do Critério de Informação de Bayes (*BIC - Bayes Information Criterion*) para estimar o número de componentes com o algoritmo da máxima verossimilhança. Além disso, é usado o método da decomposição de valores singulares (*SVD - Singular Value Decomposition*) para a redução de dimensão, e o algoritmo BFGS (Broyden, Fletcher, Goldfarb and Shanno), para otimização na inversão da matriz Hessiana (HANSEN, 2001; NIELSEN, 2001).

Dessa forma, os coeficientes ICA obtidos, a serem utilizados pelos classificadores, permitem uma outra redução de dimensionalidade, e efetuam a projeção das silhuetas em um auto-espço que mantenha a maior independência estatística entre as classes.

#### **4.7.2 O treinamento e teste dos classificadores**

Com a finalidade de se efetuar o treinamento e teste dos classificadores, para as 6 classes em estudo foram considerados os 18 maiores coeficientes independentes, que irão compor o novo vetor de características, fazendo com que a nova matriz de características, contendo os vetores de coeficientes que melhor discriminam as classes, assumisse a dimensão 42x18.

Nesta etapa, o classificador convencional K-vizinhos mais próximos (KNN), com similaridade co-senoidal, foi também utilizado para permitir comparar os resultados com os do classificador “SVM” (CANU; GRANDVALET; RAKOTOMAMONJY, 2003).

Em ambos os classificadores, a configuração de “validação cruzada”, já descrita anteriormente, foi utilizada para permitir a obtenção dos “escores” de identificação das 42 seqüências. Neste caso, o conjunto de treinamento foi formado por 41 dos vetores-silhuetas das seqüências, sendo que sempre um, de teste, fica de fora e se alterna desde o primeiro até o último para cada ciclo do processo de treinamento-validação.

A Figura 23 apresenta, para as 42 seqüências de vídeo, as duas curvas referentes ao desempenho do classificador SVM e do classificador KNN. Neste caso, os percentuais de acertos, para os dois classificadores, são obtidos a partir dos 18 maiores coeficientes independentes obtidos via método ICA sobre um número de amostras que varia em função da taxa de redução dada pela proporção de variâncias  $r_m$  desde 0 ate 10%, conforme apresentado anteriormente na Tabela 7.

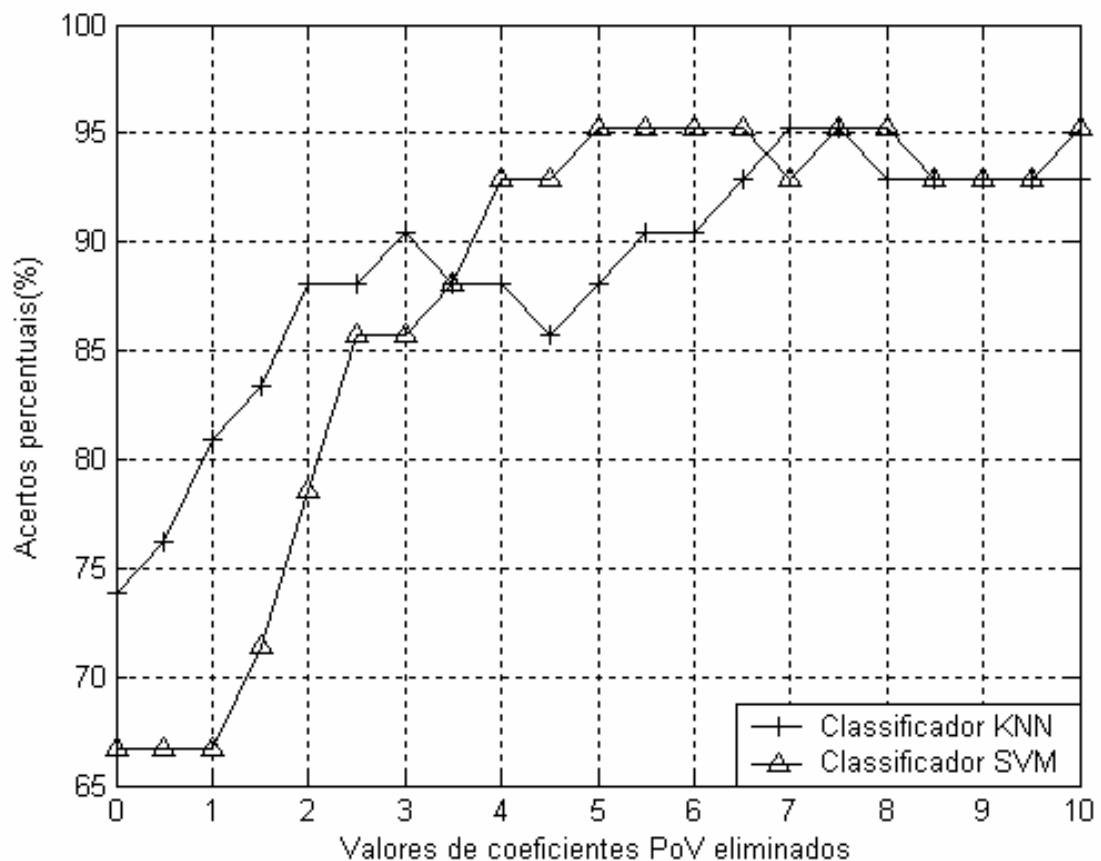


Figura 23 - Desempenho KNN x SVM para os 18 coeficientes independentes

Os resultados apresentados na Figura 23 para os dois classificadores mostram que o desempenho do classificador SVM, ao se considerar os 18 coeficientes independentes obtidos na faixa de redução percentual de  $r_m$  variando desde 3,5 a 10, foi ligeiramente superior ao do classificador K-NN e apresentou uma taxa percentual de acertos superior a 92,86.

Para permitir uma análise mais detalhada dos resultados, a Tabela 8 apresenta, na forma de uma matriz de confusão, os resultados obtidos pelo classificador “SVM” quando se aplica uma taxa de redução na proporção de variâncias de 6,5%. Aqui, foi aplicado o protocolo de validação-cruzada sobre as 7 seqüências das 6 classes da base de dados UCSD, totalizando 42 vetores de 18 coeficientes independentes cada.

Tabela 8 - Matriz de confusão para o classificador SVM

CLASSES		SAÍDA DO CLASSIFICADOR SVM						PRECISÃO (%)
		1	2	3	4	5	6	
R E A I S	1	6	0	1	0	0	0	85,71
	2	0	7	0	0	0	0	100,0
	3	0	0	7	0	0	0	100,0
	4	0	0	0	6	1	0	85,71
	5	0	0	0	0	7	0	100,0
	6	0	0	0	0	0	7	100,0
CONFIANÇA		100,0	100,0	87,50	100,0	87,50	100,0	Total: 95,24

Considerando-se os resultados apresentados na Tabela 8, pode-se observar que o percentual de acertos de 95,24 % por seqüência de silhuetas está de acordo com aqueles obtidos por distintos métodos de extração e seleção de características. Além disso, eles apresentam resultados com desempenho comparável àqueles trabalhos que usam outros algoritmos de classificação, como os apresentados por Phillips et al. (2002) e Wang et al. (2003b), o que nos incentiva a aprofundar as pesquisas da metodologia proposta.

#### 4.8 Implementação com abordagem PoV e classificador RBF

Nesta implementação, propomos uma arquitetura simples, conforme apresenta a Figura 24, cuja modelagem procura aproximar-se do raciocínio humano na tomada de decisões. Neste caso, consideramos estudos sugerindo que o ser humano é capaz de reconhecer seus semelhantes comparando suas faces e silhuetas com uma imagem média armazenada no cérebro, e adaptamos a técnica capaz de representar em uma única imagem os índices de energia das silhuetas para obter as informações estatísticas mais relevantes contidas na biomecânica das silhuetas de um ciclo (PHILLIPS et al., 2002).

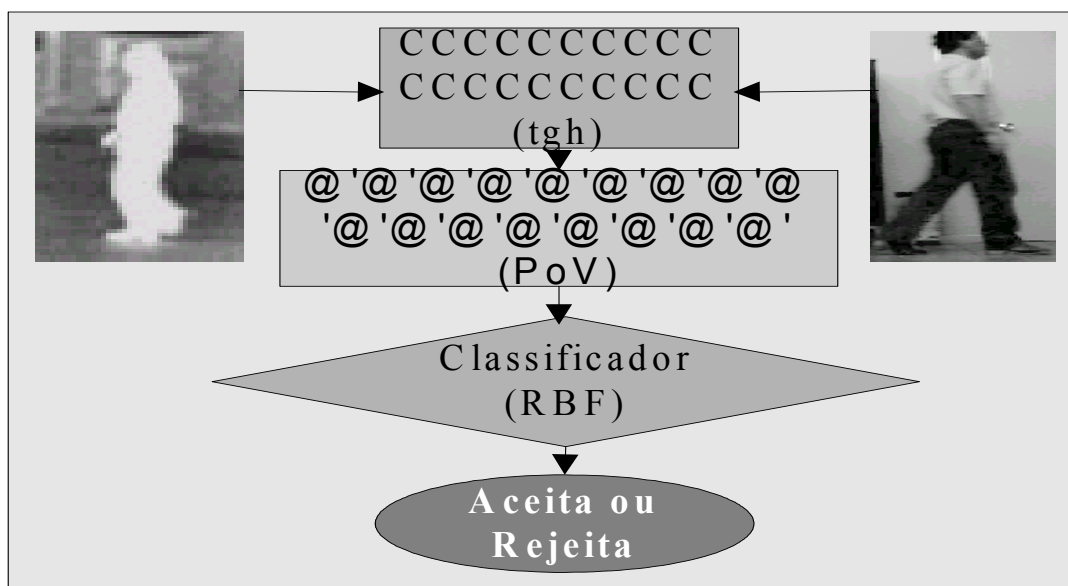


Figura 24 - Arquitetura proposta

A nossa proposta utiliza-se das silhuetas de seqüências de vídeo do caminhar humano que são representadas em uma única imagem, após aplicar-se a técnica da Energia das Silhuetas do Caminhar (ESC), adaptada de *Motion-Energy-Image/Motion-History-Image (MEI/MHI)* (BOBICK; DAVIS, 2001). Esta técnica necessita de, pelo menos, um ciclo do caminhar nas seqüências de vídeo capturadas, para que seja possível o reconhecimento humano automático pela forma de caminhar. Além disso, ela pode ser aplicada tanto às seqüências de vídeo obtidas com câmaras no espectro visível quanto no infravermelho, além de representar a seqüência do caminhar em uma única imagem, e ainda preservam as informações temporais do ciclo (HAN; BHANU, 2005).

Além disso, aplicamos duas técnicas estatísticas sobre os vetores de características básicos: a tangente hiperbólica (tgh), para padronizar as características do conjunto de vetores de características; e a proporção de variâncias (PoV) que, além de efetuar a seleção das características de maior poder discriminatório entre as classes, efetua, ainda, a redução de dimensionalidade do conjunto de vetores (INDOVINA et al., 2003).

Concluimos nosso trabalho utilizando dois classificadores, sendo um baseado em distâncias Euclidianas e, outro, em uma rede neural com funções radiais de base (RBF), onde apresentamos e discutimos os resultados obtidos na identificação pessoal.

#### 4.8.1 A Base de Dados utilizada

Foi utilizada uma base de dados de domínio público disponibilizada pelo Georgia Institute of Technology para a validação da proposta, conforme alguns exemplos apresentados na Figura 25 (GATECH, 2006). Ela contém 18 pessoas caminhando de perfil diante de uma câmara em um ambiente com iluminação semi-controlada, gerando um total de 6 seqüências de vídeo, no formato AVI, para cada pessoa.

A captura das seqüências foi realizada à razão de 33 quadros por segundo, e cada seqüência, dependendo principalmente das dimensões do corpo da pessoa, tem entre 65 e 120 quadros, com resolução de 320 x 240 pixels cada.



Figura 25 – Exemplos de quadros de seqüências da base de dados utilizada

A obtenção dos resultados inicia-se com a etapa de pré-processamento das seqüências de vídeo e da aplicação da técnica ESC. Em seguida, para cada etapa de aplicação do coeficiente de redução de proporção das variâncias (PoV), efetua-se a correspondente classificação de todos os vetores que emite um rótulo da identidade pessoal.

#### 4.8.2 As Etapas de pré-processamento

O pré-processamento inicia-se com a captura das seqüências de vídeo do caminhar. A seguir, são extraídos, em tons de cinza, os quadros da seqüência de vídeo e o



fundo do cenário. As silhuetas cheias, conforme apresenta a Figura 26(a), são obtidas aplicando-se um algoritmo de subtração que compensa o brilho dos pixels ao correlacionar-se o quadro atual e o fundo do cenário. Além desse, são aplicados algoritmos que delimitam a janela dinâmica, executam a filtragem e a “binarização”, bem como a conexão dos componentes vizinhos da silhueta interna (SALOMÃO; ALMEIDA; SALLES, 2006b).

#### 4.8.3 A Energia das silhuetas do caminhar

Dado um conjunto de silhuetas com dois níveis de cinza  $S_t(x,y)$  de um ciclo contendo  $N$  silhuetas, conforme apresenta a Figura 26(a), o algoritmo para se obter a imagem ESC da silhueta com seus níveis de energia em tons de cinza, conforme apresenta a Figura 26(b), é implementado através da expressão

$$ESC(x, y) = \frac{1}{N} \sum_{t=1}^N S_t(x, y), \quad (40)$$

onde  $x$  e  $y$  são as coordenadas da imagem-silhueta e  $t$  é o número do quadro na seqüência de silhuetas do ciclo em um dado instante de tempo.

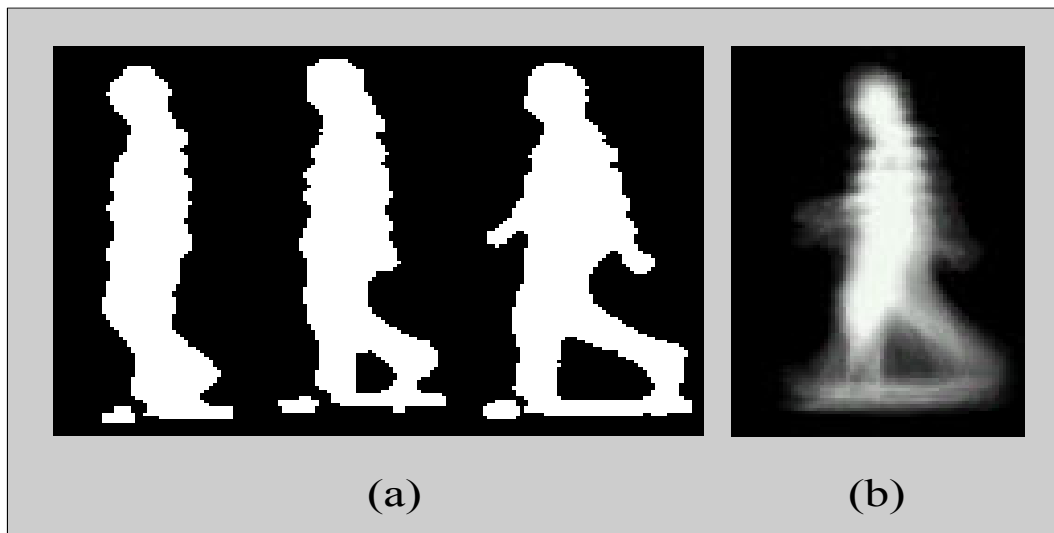


Figura 26 – Exemplos de silhuetas de um ciclo e imagem ESC

A partir da imagem da Energia das Silhuetas do Caminhar (ESC) e através da varredura de cima para baixo de todos os pixels em ordem lexicográfica das linhas da imagem, o algoritmo concatena as informações da posição e níveis de energia de

cada coeficiente na imagem para constituir os 108 vetores básicos de características, contendo cada um 5550 amostras.

#### 4.8.4 Os Classificadores utilizados: Euclidiano e RBF

O treinamento e teste dos vetores de características são feitos através da modalidade extrema de validação-cruzada do tipo “deixe um de fora”. Neste caso, em cada ciclo de execução, o algoritmo retira sequencialmente um vetor-teste do conjunto total de vetores e, em seguida, sobre os N-1 vetores restantes, processa a classificação do vetor retirado e emite um rótulo de identidade. Tal procedimento se repete exaustivamente, prosseguindo até o último vetor do conjunto quando, então, é feito o cálculo da taxa de acertos.

Nessa abordagem, para fins de comparação de desempenho, um classificador clássico baseado em distâncias Euclidianas é utilizado simultaneamente com outro classificador RBF, que têm a função de aceitar ou rejeitar o acesso da pessoa.

A Figura 27 apresenta, para os classificadores Euclidiano e RBF, os percentuais de acertos com coeficientes de redução PoV variando entre 0 e 0,4, o que corresponde a um total de 5550 e de 1250 amostras, respectivamente.

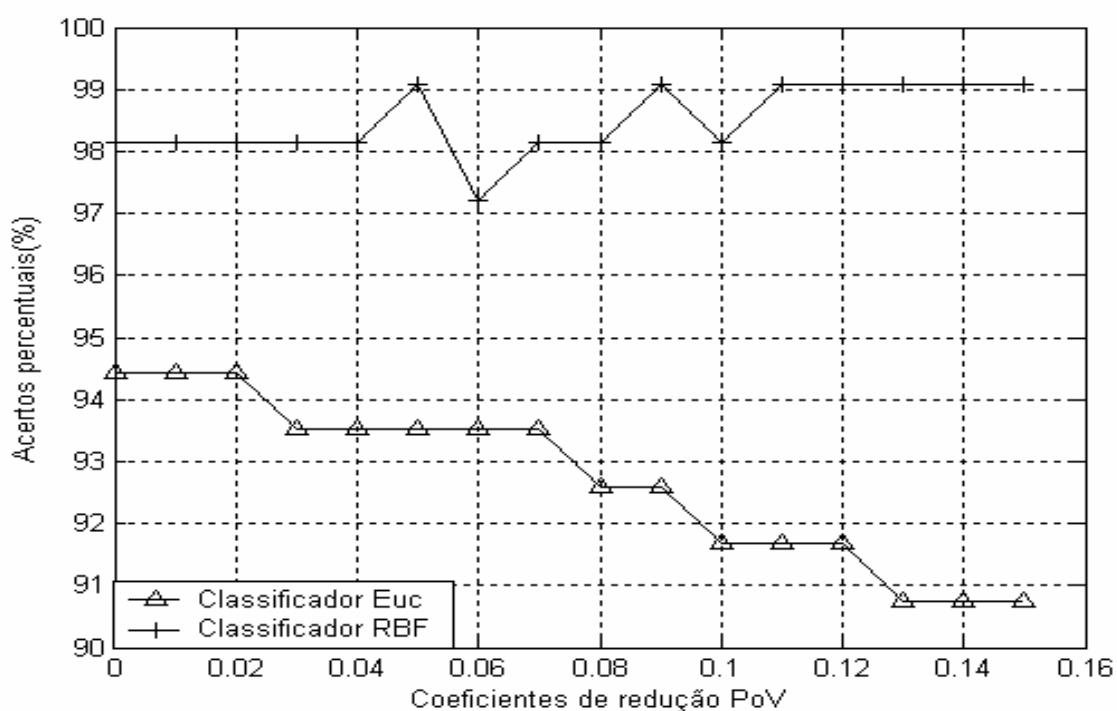


Figura 27 - Resultados obtidos

Considerando os vetores de características dentro da faixa de 0 a 0,04 dos coeficientes PoV apresentados na Figura 27, o que ainda corresponde a um grande número de amostras, implicando em um baixo desempenho computacional devido a um maior custo de processamento, observa-se que ocorre o maior desempenho do classificador Euclidiano. A partir desse ponto, para valores de coeficientes PoV entre 0,04 até 0,40, observa-se um excelente desempenho do classificador RBF e uma queda gradativa no desempenho do classificador Euclidiano. Aqui, deve-se ainda levar em conta que foi utilizado o Toolbox de Redes Neurais do Matlab®, cujo algoritmo RBF não é dos mais eficientes.

Observamos, ainda, através das curvas da Figura 27, que as taxas de acertos de 99% são mais estáveis e confiáveis para o classificador RBF, e ocorrem quando os coeficientes PoV assumem valores compreendidos entre 0,12 a 0,20. Aqui, as vantagens do classificador aparecem também em função do baixo custo computacional no processo de treinamento e da eficácia na etapa de classificação, uma vez que as amostras são reduzidas para um total de 1250 por vetor de características.

#### **4.9 Considerações finais**

Neste capítulo, iniciamos nosso estudo com implementações de sistemas automáticos de reconhecimento pessoal baseando-se simplesmente em silhuetas extraídas de seqüências de vídeo, avançamos para um estudo baseando-se em vetores de características constituídos com todos os quadros da seqüência e finalizamos com a técnica da energia das silhuetas de uma seqüência na qual o vetor de características é extraído de uma única imagem. Nesse estudo, nosso objetivo principal, além de avaliar o desempenho de sistemas de reconhecimento humano baseados nas características da forma de caminhar, é o de encontrar, em algumas das técnicas utilizadas, indícios que facilitem a fusão de seqüências da forma de caminhar com imagens faciais.

Na primeira implementação, aplicamos algumas técnicas que envolvem a fase de pré-processamento para a obtenção das características biométricas, bem como as transformadas wavelets e a combinação dela com a técnica PCA na seleção das

mesmas. Propomos e avaliamos o desempenho dos classificadores Neural e K-vizinhos mais próximos.

Na segunda implementação, apresentamos um breve estudo de um dos métodos utilizados no reconhecimento humano, com o objetivo principal de avaliar o desempenho do classificador baseado nas máquinas de vetores suporte quando a seleção de características é feita através das transformadas Wavelets. Neste caso, o classificador SVM, além de apresentar um melhor desempenho em relação ao K-NN, ainda tem a vantagem de permitir armazenar os parâmetros de treinamento obtidos a partir da base de dados ao agrupar as características mais correlacionadas e rotular as diversas classes de acordo com esse critério, possibilitando, assim, um menor tempo de processamento na classificação dos vetores de teste.

Nas duas implementações, os resultados obtidos estão compatíveis com aqueles apresentados por pesquisadores que utilizaram outras técnicas na seleção e na classificação dos vetores de características. Em especial, é interessante compará-los com os trabalhos desenvolvidos por Wang et al, (2003b) e Hayfron-Acquah, Nixon e Carter (2003), que utilizaram técnicas tradicionais em reconhecimento de padrões ao combinarem a Análise de Componentes Principais com o classificador K-vizinhos mais próximos.

Na terceira implementação, foi utilizada a abordagem PoV/ICA na seleção das características associadas com um classificador neural do tipo RBF. Aplicamos a técnica da Proporção de Variâncias em conjunto com o método de Análise de Componentes Independentes na seleção dessas características e avaliamos o desempenho do classificador SVM, que se mostrou, através da comparação dos resultados obtidos com outras técnicas utilizadas, como uma ferramenta adequada ao desafio do reconhecimento humano pelas características biomecânicas da sua forma de caminhar.

Na última implementação, utilizamos a técnica da energia das silhuetas de um ciclo do caminhar (ESC) associada com a técnica PoV na fase de extração e seleção das características em conjunto com um classificador neural do tipo RBF. Neste caso, além do comprovado desempenho obtido pelo classificador, observamos que a imagem contendo os coeficientes de energia da dinâmica do caminhar, quando

associada com a técnica PoV, pode contribuir e facilitar a fusão no nível de características das seqüências de vídeo da forma de caminhar com as imagens faciais.

Nas duas últimas implementações, os resultados alcançados mostraram os bons desempenhos das técnicas de seleção e dos classificadores escolhidos e foram superiores ou estão de acordo com aqueles obtidos por distintos métodos de extração e seleção de características e que fazem uso de outros algoritmos de classificação (CUNADO et al., 1999); (BOULGOURIS; HATZINAKOS; PLATANIOTIS, 2005); (HAN; BHANU, 2006); (FOSTER; NIXON; PRÜGEL-BENNETT, 2003) (WANG et al., 2004), o que nos garante a viabilidade da proposta por ter se mostrado confiável e adequada ao desafio do reconhecimento humano.

No próximo Capítulo, conforme nossa proposta inicial, são avaliadas algumas modalidades da fusão dos dois sistemas biométricos face e forma de caminhar. Neste caso, é avaliada a aplicabilidade dessas e de outras técnicas de seleção e de classificação em problemas de reconhecimento pessoal que permitam estabelecer a fusão das características biométricas faciais e da forma de caminhar.

# Capítulo 5

## **A Identificação Pessoal por meio da Fusão Face-Forma de Caminhar**

Um sistema biométrico multimodal procura integrar as informações de diversas fontes biométricas para compensar as limitações no desempenho de cada sistema individual. Dessa forma, procura-se utilizar a fusão de indicadores biométricos distintos, consistindo em combinar duas ou mais modalidades biométricas para se obter um melhor desempenho no reconhecimento. A fusão é possível em vários níveis ao longo do processo, podendo ser aplicada desde a fase em que as características são obtidas diretamente dos sensores até à fase final, ou seja, no nível de resultados dos classificadores.

Nossa proposta permite estabelecer a fusão dos dois indicadores biométricos formados por imagens faciais e silhuetas da forma de caminhar. As imagens podem ser obtidas a partir de filmes originados de câmaras operando nos espectros visível, infravermelho ou termográfico. No caso das imagens faciais termográficas, um estudo com bons resultados de classificação foi feito com uma base de dados de domínio público, conforme apresentado em Salomão, Almeida e Salles (2006a). Entretanto, em consequência do elevado custo e das dificuldades em se adquirir os equipamentos necessários à captura das imagens, nosso foco ficou voltado para imagens capturadas no espectro visível.

Neste Capítulo, serão descritas e apresentadas duas arquiteturas de fusão com seus respectivos algoritmos a serem empregados em um sistema bi-modal de reconhecimento pessoal. As propostas estão relacionadas à fusão no nível das características de imagens faciais e de silhuetas de seqüências de vídeo da forma de caminhar capturadas no espectro visível.

### **5.1 As diversas formas de se estabelecer a fusão biométrica**

O problema do reconhecimento pessoal através da fusão das características biométricas pode assumir as seguintes formas (ROSS; JAIN, 2004):

- fusão múltipla de classificadores heterogêneos atuando sobre uma única característica biométrica. Neste caso, os rótulos de classes emitidos pelos múltiplos classificadores são avaliados em um sistema de voto ponderado ou majoritário;
- fusão de múltiplos escores obtidos sobre uma única característica biométrica. Neste caso, os diversos escores das medidas que mais se aproximam dos modelos na base de dados gerados por múltiplas estratégias (na mesma representação) são combinados;
- fusão de sistemas biométricos contendo múltiplas características envolvendo diferentes modalidades biométricas de entrada, que será abordada neste trabalho.

### **5.2 A fusão de sistemas biométricos com múltiplas fontes de informações**

Novas pesquisas de identificação pessoal, utilizando-se da composição da forma de caminhar e de imagens faciais, estão em foco hoje em alguns Institutos e Universidades do mundo. Elas são necessárias, uma vez que, apesar da vantagem em permitir a identificação à distância, a confiabilidade de somente um dos dois sistemas biométricos é considerada baixa em comparação com outros sistemas biométricos, como os baseados em impressões digitais, íris e retina (JAIN; ROSS, 2004).

O aumento das possibilidades de identificação à distância, em ambientes com pouca ou nenhuma luminosidade, exige que as técnicas propostas sejam flexíveis e capazes de processarem as informações obtidas de imagens dos espectros visível e infravermelho. Atualmente, há poucas publicações nesta área, e, entre elas, destacam-se os trabalhos de Kale, Roy-Chowdhury e Chellappa (2002), Shakhnarovich, Lee e Darrel (2001) e Shakhnarovich e Darrel (2002).

A Figura 28 apresenta um diagrama de um sistema biométrico e comportamental, na modalidade bi-modal, constituído de imagens faciais e silhuetas de seqüências de vídeo do caminhar que podem ser obtidas tanto no espectro visível quanto no infravermelho ou termográfico.

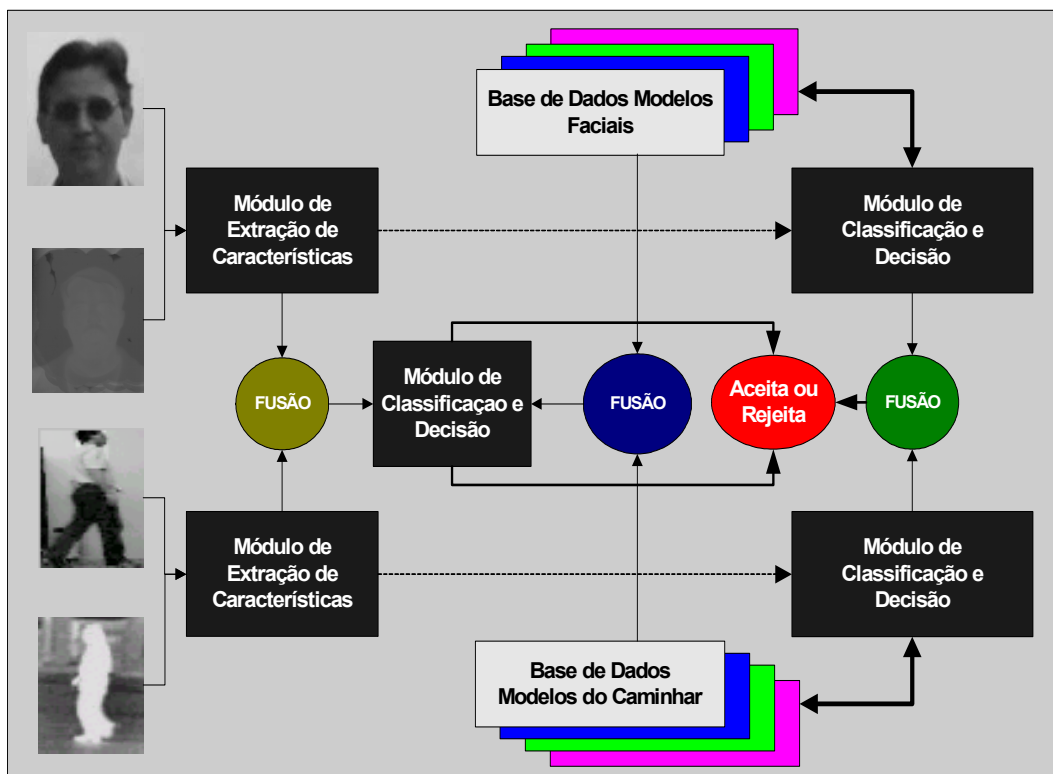


Figura 28 – Estrutura de um sistema bi-modal: módulos de fusão face-forma de caminhar

Neste caso, o objetivo é o de ilustrar os três níveis de fusão de informações apresentados em Jain e Ross (2004) para sistemas multi-modais e que são descritos a seguir:

- **fusão no nível de extração das características** – Neste caso, o dado obtido de cada sensor é usado para compor o vetor de características. As características extraídas de um dos sistemas biométricos são, inicialmente, tratadas de forma independente daquelas extraídas do outro e, em seguida, os dois vetores são combinados para se obter um novo vetor, no mesmo ou em outro hiper-espaco e com a mesma ou diferente dimensão dos originais;
- **fusão no nível da combinação de escores** – Cada sistema providencia uma relação de escores que indica o quanto um vetor de teste está próximo dos vetores de características armazenados no modelo da base de dados, de



acordo com uma função de similaridade. Estes escores são combinados para assegurar a identidade da pessoa. Diversas técnicas podem ser empregadas nesse nível: uma delas, a regressão logística, que procura minimizar a FRR para uma dada FAR, permite combinar os escores;

- **fusão no nível de decisão** – Cada sensor atua separadamente na captura dos múltiplos dados biométricos, e os resultados gerados pelos classificadores sobre os vetores de características são individualmente classificados em duas classes, do tipo aceita ou rejeita. A decisão final é, então, tomada em função de um esquema de voto majoritário.

Uma melhor eficiência de um sistema multimodal pode ser conseguida ao se adotarem combinações de regras em múltiplos estágios. Neste caso, em função de uma métrica escolhida, ao se apresentar um indivíduo teste, e considerando toda a base de dados, um determinado número de indivíduos é identificado como pertencente a uma classe por um classificador preliminar. Em seguida, esse número reduzido de possíveis candidatos passa novamente pelo crivo de um classificador mais preciso, que irá, utilizando uma métrica que possa ou não levar em conta os resultados preliminares, estabelecer a identidade a ser atribuída ao indivíduo.

Na próxima seção iremos apresentar as propostas para o cenário, arquiteturas e algoritmos que permitirão efetuar a fusão das características faciais e da forma de caminhar na implementação de sistemas de reconhecimento pessoal.

### **5.3 As propostas para os esquemas de fusão a serem utilizadas**

Para fins de avaliação e análise de desempenho de classificadores, optamos por estabelecer a fusão de duas formas distintas: uma, considerando a fusão no nível de decisão, isto é, fundindo os escores dos classificadores e combinando-os em um esquema de voto simples; e outra, considerando o nível das características, isto é, fundindo as características faciais e da forma de caminhar.

Neste trabalho, para as duas propostas de implementação de fusão face-forma de caminhar, as características faciais e da forma de caminhar são normalizados pela técnica da tangente hiperbólica, para concentrar suas representações. A seleção de características, com a respectiva redução de dimensionalidade, é processada

através da aplicação da técnica da proporção de variâncias. Para a etapa de classificação, para fins de análise de desempenho e comparação dos resultados, utilizam-se os classificadores por distância Euclidiana e classificador neural com função de base radial.

### 5.3.1 As bases de dados de domínio público

A validação da nossa proposta foi feita utilizando-se de duas bases de dados de domínio público, sendo que uma delas é constituída de imagens faciais.

A primeira base de dados, para a qual alguns exemplares são mostrados na Figura 29(a), mostra 6 seqüências de vídeo no formato AVI para cada uma das 20 pessoas caminhando de perfil em relação à câmara (GATECH, 2006). Neste caso, cada seqüência de vídeo é capturada à razão de 33 quadros por segundo e tem, aproximadamente, entre 65 e 120 quadros na resolução de 320 x 240 pixels a serem extraídos.

A segunda base de dados usada, conforme os exemplares de imagens faciais da Figura 29(b), contém 40 pessoas com um total de 10 imagens faciais visíveis em tons de cinza para cada uma, apresentadas na resolução 92 x 112 pixels (OLIVETTI, 2005).



Figura 29 – Exemplos de quadros das seqüências de vídeo e imagens faciais utilizadas

Em nossa implementação, de cada base de dados foram consideradas 18 pessoas e nossos algoritmos utilizaram sempre, para compor os vetores de características, um total de 6 seqüências de vídeo e de igual número das imagens faciais por pessoa.

Por serem distintas as duas bases de dados, associamos os rótulos de identidade das seqüências de vídeo do caminhar de uma base com as imagens faciais da outra. Além disso, por fazer uso das silhuetas cheias, embora não seja esse o foco atual da pesquisa, nosso algoritmo pode ser aplicado indiferentemente para imagens obtidas em qualquer um dos espectros, uma vez que a obtenção do vetor de características segue as mesmas etapas e faz uso das mesmas técnicas e ferramentas empregadas em nosso trabalho.

### **5.3.2 As Etapas de pré-processamento**

As duas abordagens a serem implementadas baseiam-se em arquitetura bi-modal face-forma de caminhar. Sendo assim, a fase de pré-processamento inicia-se com a captura das imagens faciais e das seqüências de vídeo do caminhar, que irão compor as diversas bases de dados. A seguir, são extraídos, em tons de cinza, os quadros da seqüência de vídeo e o fundo do cenário, que, em conjunto, permitirão obter as silhuetas de um ciclo do caminhar. Assim, a partir das imagens faciais em tons de cinza, e da imagem com os coeficientes da Energia das Silhuetas do Caminhar (ESC), utiliza-se a técnica de varredura para se obter as características contidas nos pixels que, constituirão para cada imagem e para cada seqüência, os diversos vetores de características.

O vetor de características das imagens faciais pode ser extraído de uma imagem com origem em um conjunto de poses discretas ou originadas de uma seqüência de vídeo. Assim, o vetor de características foi obtido varrendo-se as linhas e colunas da imagem e formando um vetor contendo os níveis de cinza correspondentes a cada pixel da imagem, conforme a Figura 30(a).

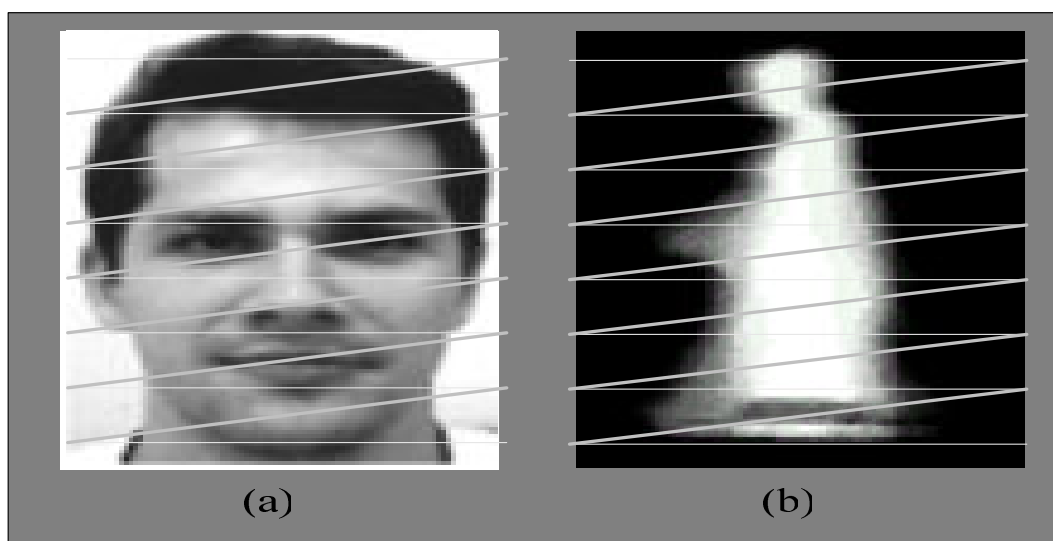


Figura 30 - Obtenção dos vetores de características da face e da silhueta

O vetor de características do caminhar exige primeiro a obtenção do fundo do cenário e da sequência de silhuetas cheias, em dois níveis de cinza (SALOMÃO; ALMEIDA; SALLES, 2005b). Em seguida, após ser obtida a imagem da Energia das Silhuetas do Caminhar (ESC), conforme sugere a Figura 30(b), é aplicado o algoritmo que concatena os coeficientes das linhas da matriz-imagem e capta as informações da posição e dos níveis de energia de cada pixel, dando origem ao vetor de características básicas.

Ainda para as duas implementações a serem descritas, o treinamento e teste dos vetores de características das faces e da forma de caminhar foram feitos utilizando-se classificadores com a arquitetura de validação cruzada.

#### 5.4 Implementação da fusão no nível de decisão

A arquitetura proposta sugere a fusão bi-modal de forma hierárquica que utiliza as imagens faciais e as seqüências de vídeo da forma humana de caminhar no reconhecimento pessoal. A fusão proposta ocorre no nível de decisão, onde um esquema de voto permite uma pré-identificação do caminhante por um dos classificadores.

Assim, a forma de caminhar, além de permitir a identificação prévia da pessoa, emite ainda um escore de classificação que funciona como um filtro que irá permitir a fusão com os resultados do classificador das imagens faciais.

Apesar da flexibilidade do cenário permitir o uso de várias arquiteturas na fusão face-forma de caminhar, usamos os escores das características da forma de caminhar como uma espécie de filtro, onde o classificador trabalha de forma mais relaxada, selecionando alguns vetores de características e associando-os previamente a possíveis identidades.

Na etapa seguinte, outro classificador, mais preciso, utilizando-se dos vetores de características das imagens faciais, compara o seu resultado com os escores obtidos na etapa anterior, concluindo pela aceitação ou pela rejeição final da pessoa, como mostra a Figura 31.

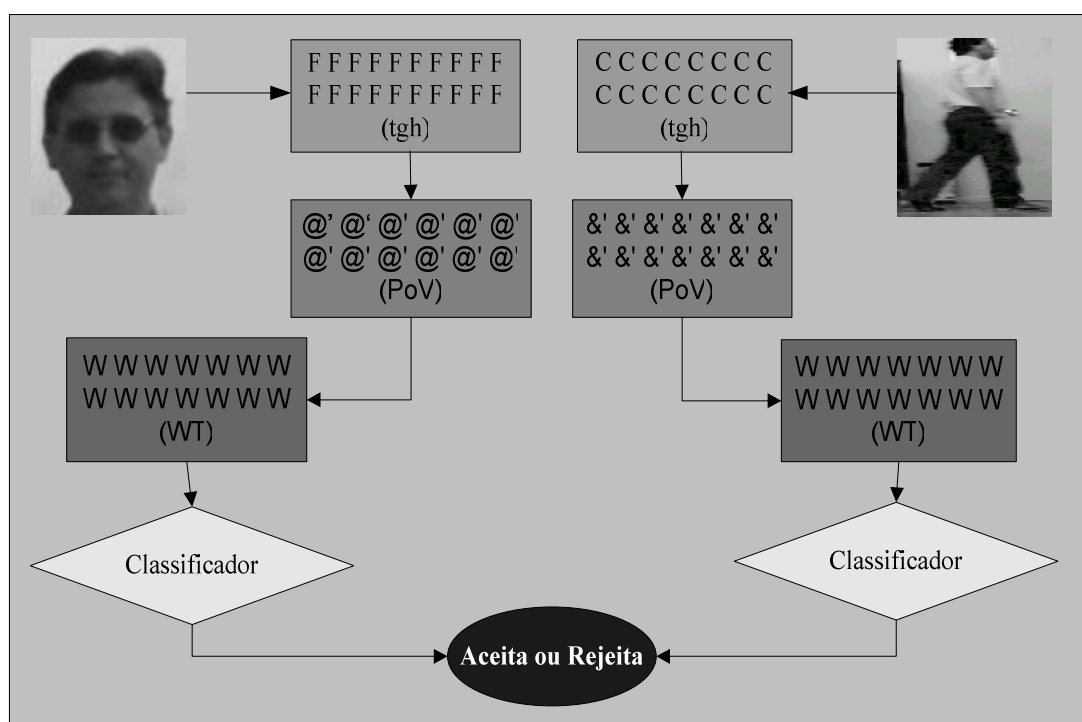


Figura 31 – Arquitetura com o esquema de fusão no nível de decisão

Nessa abordagem, para maior confiabilidade do processo, um primeiro classificador, baseado em distâncias euclidianas é utilizado para se obter as mais prováveis classes das seqüências do caminhar. Em seguida, apesar de exigir um maior custo computacional, outro classificador do tipo K-vizinhos mais próximos, por permitir ajuste para se obter um melhor desempenho, é utilizado para encontrar a classe do vetor associada com a face da pessoa que caminhou diante da câmara.

Na etapa seguinte, um critério de decisão que considera o rótulo da face e até as três primeiras identidades associadas com as menores distâncias do caminhar, aceita ou rejeita o acesso da pessoa. Aqui, o voto vencedor executa uma decisão do tipo tudo-ou-nada. Isto é, um dos três padrões teste da base de dados do caminhar que formaram o escore de identidades deve coincidir a sua classe contendo o padrão teste da base de dados faciais para se ter uma identidade genuína. Caso contrário, este padrão não terá sua identidade validada e a pessoa terá seu acesso rejeitado, sendo classificada como uma impostora.

#### **5.4.1 A etapa de extração e seleção de características**

O procedimento de extração e seleção de características segue as mesmas etapas descritas anteriormente, onde os vetores de características da forma de caminhar e das faces são montados separadamente. A técnica da Energia das Silhuetas do Caminhar é usada para extrair o vetor de características das seqüências. Em seguida, aplica-se a técnica estatística de proporção de variâncias (PoV) para efetuar a primeira etapa de seleção de características, assim como iniciar o processo da redução de dimensionalidade sobre os vetores de características, e a transformada Wavelet (WT), que permite gerar os coeficientes de maior energia para a fase de classificação.

#### **5.4.2 O procedimento de classificação**

Na etapa final, um classificador baseado em distância atua sobre os vetores da forma de caminhar e emite um escore das dez classes mais prováveis em ordem decrescente. Esta informação é armazenada e, posteriormente, é comparada com o resultado de outro classificador “K-vizinhos mais próximos”, este atuando sobre os vetores das imagens faciais.

O algoritmo final de tomada de decisão, baseado em um esquema de voto majoritário, toma a classe rotulada para a imagem facial e leva em conta a coincidência de identidade de uma das três primeiras classes emitidas pelo escore do classificador da forma de caminhar. Assim, apesar da taxa de acertos assumir valores intermediários, a proposta permite implementar um sistema de reconhecimento pessoal com uma classificação prévia, por meio da forma de

caminhar, fazendo com que ela atue como um filtro para a identificação final do indivíduo.

### 5.4.3 Análise dos resultados

A Figura 32 apresenta, para as 108 seqüências de vídeo do caminhar e as correspondentes classes de imagens faciais, as três curvas com os acertos percentuais dos classificadores em função do coeficiente de redução PoV. Observou-se que, considerando a eliminação de coeficientes menos discriminatórios sobre os vetores de características em uma faixa de 0 a 0,15 dos valores de  $r_m$  normalizados, é possível obter desempenhos que permitem uma melhor avaliação dos resultados.

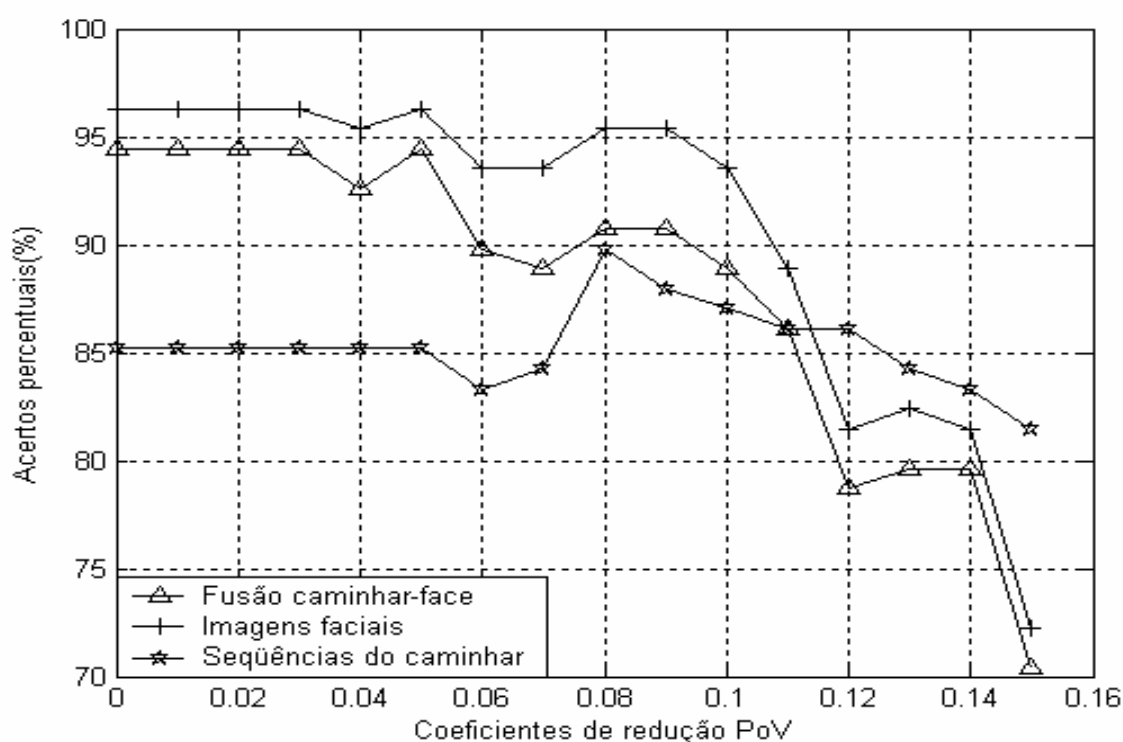


Figura 32 - Resultados obtidos: fusão no nível de decisão

Observamos, através das curvas da Figura 32, que a taxa de acertos da fusão no nível de decisão é menor que a dos sistemas biométricos individuais. Isto ocorre em consequência do critério de decisão adotado baseado no conectivo E (AND), onde, para tornar o sistema mais confiável, a identidade só é validada se ambos os





A dimensão dos vetores de características de ambos os sensores biométricos é ainda normalizada, tomando-se como referência a menor dimensão entre os vetores de características faciais ou da forma de caminhar e aplicando-se a técnica PoV. Este procedimento de seleção permite, ainda, reduzir a dimensionalidade do problema ao estipular-se o número de coeficientes a serem considerados dentre as características que melhor discriminam as variâncias dentro e entre as classes.

A partir do ponto em que as bases de dados estão normalizadas e padronizadas, são possíveis estabelecer vários esquemas de fusão das características, entretanto, em nosso trabalho para simplificação dos algoritmos, foi executado um esquema simples baseado na média da soma das características. Na etapa seguinte, aplica-se a técnica PoV, agora objetivando, simultaneamente, a redução de dimensionalidade e um maior poder discriminatório dos coeficientes dos vetores de características a serem utilizados na fase de classificação.

Para a etapa de classificação das imagens faciais, das silhuetas extraídas das seqüências do caminhar, bem como para a fusão de ambas, em consequência do bom resultado alcançado em outra implementação para a forma de caminhar apresentada no capítulo anterior, optamos por um classificador neural com funções de base radial (RBF). Neste caso, sua implementação é tal que o número de neurônios da camada escondida fica livre, podendo alcançar um valor menor ou igual ao número de classes.

### **5.5.1 Resultados obtidos com as bases de dados ORL/GATECH**

A Tabela 9 apresenta as taxas de acerto obtidas para as faces ( $Fa$ ), forma de caminhar ( $Ca$ ) e da fusão de ambas ( $Fu$ ), em função da taxa normalizada de redução PoV ( $r_m$ ), do número de coeficientes dos vetores de características ( $Cf$ ) e do tempo de processamento ( $Tp$ ) em segundos necessário à normalização, treinamento e classificação de toda a base de dados.

Os resultados foram obtidos considerando-se uma taxa de eliminação de coeficientes menos discriminatórios sobre os vetores de características dentro da faixa de 0,00 a 0,20 dos valores de  $r_m$  normalizados.

Tabela 9 – Taxas de acertos: faces ( $Fa$ ), caminhar ( $Ca$ ) e a sua fusão ( $Fu$ ) em função da taxa PoV ( $r_m$ ), da dimensão do vetor de características ( $Cf$ ) e do tempo de processamento ( $Tp$ )

Fusão Face-Forma de Caminhar no Nível de Características																	
$r_m$	0,00	0,05	0,06	0,07	0,08	0,09	0,10	0,11	0,12	0,13	0,14	0,15	0,16	0,17	0,18	0,19	0,20
$Cf$	5551	2370	1856	1294	938	626	427	336	296	269	248	222	193	179	157	150	141
$Tp$	3023	1235	938	705	482	346	235	160	145	130	125	120	112	105	95	90	87
$Fa$	2,8	60,2	80,6	96,3	97,2	97,2	97,2	96,3	95,4	96,3	92,6	92,6	83,3	72,2	69,4	62,9	64,8
$Ca$	92,6	99,0	98,1	98,1	99,0	95,4	83,3	90,8	89,9	88,9	90,7	90,7	87,9	85,2	82,4	82,4	83,3
$Fu$	76,8	99,1	100	99,1	99,1	100	100	100	100	98,1	99,1	98,1	98,1	96,3	95,4	91,7	89,8

Devido à não linearidade de  $r_m$ , observamos que o número de coeficientes dos vetores de características não se altera de forma linear. Neste caso, o tempo de processamento ( $Tp$ ) cresce de forma exponencial com o aumento dos coeficientes do vetor, enquanto que a taxa de acertos diminui.

A Figura 34 apresenta uma síntese dos dados da Tabela 9, para as 108 seqüências, nas três curvas com os acertos percentuais do classificador RBF para as imagens faciais, as seqüências do caminhar e a fusão de ambas em função do coeficiente de redução PoV. Neste caso, quando ocorre a eliminação de coeficientes menos discriminatórios dos vetores de características em uma faixa de  $r_m$  variando de 0,06 a 0,15, é possível obter os melhores desempenhos do classificador na fusão de características.

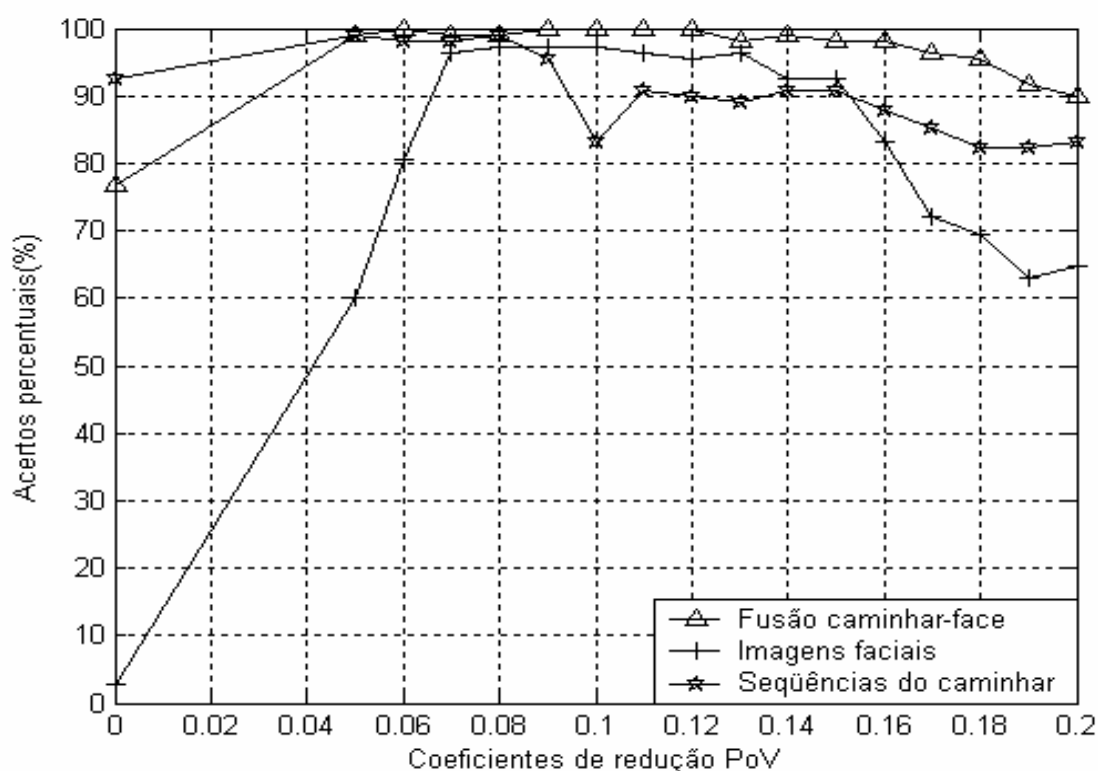


Figura 34 - Resultados obtidos: fusão no nível de características

Observamos, ainda, através das curvas da Figura 34, que as taxas de acertos para as imagens faciais para valores de  $r_m$  variando de 0,00 a 0,05 estão com valores extremamente pequenos, se comparados com as características da forma de caminhar. Uma das possíveis explicações para este fato é a baixa capacidade de discriminação oferecida pelo grande número de coeficientes dos vetores de características para o classificador RBF.

Apesar dos bons resultados obtidos, que comprovaram a nossa hipótese de que a fusão no nível de características traria um melhor desempenho de classificação ao sistema de reconhecimento pessoal, propomos obter, em dois momentos distintos, bases de dados com seqüências de vídeo e imagens faciais dos mesmos indivíduos. Sendo assim, no próximo capítulo, além de apresentarmos as respectivas bases de dados, avaliaremos o desempenho de classificação e analisaremos, apesar de ainda ser limitado o número de indivíduos, alguns fatores e taxas que refletem a robustez do sistema de reconhecimento pessoal.

## 5.6 Considerações finais

Neste Capítulo, propusemos duas arquiteturas para um sistema de reconhecimento pessoal que permitiram efetuar a fusão das características faciais e da forma de caminhar em dois níveis, uma no nível de decisão ou de classificadores e, outra, no nível das características, sendo ambas implementadas sobre bases de dados de domínio público. Nos dois casos, fizemos uso de algoritmos e técnicas de extração e seleção de coeficientes desde a fase de pré-processamento, para obter as silhuetas e a imagem contendo os coeficientes de energia da dinâmica do caminhar, até a fase final de classificação e emissão de rótulos de identidades.

Na primeira implementação, apesar dos dados pertencentes às duas modalidades não estarem disponíveis para um único conjunto de usuários, a suposição da independência mútua destes dois indicadores biométricos permite-nos coletar os dados biométricos individualmente de bases de dados não representando os mesmos indivíduos e, então, concatená-los. Esta implementação consistiu em combinar os escores obtidos dos sistemas faciais e da forma de caminhar dos indivíduos.

Para tanto, aplicamos a técnica da proporção de variâncias e da transformada Wavelet na seleção dos coeficientes e utilizamos os escores de classificação sobre a forma de caminhar como um filtro que atua na decisão final após o classificador sobre as imagens faciais. Os resultados apresentaram um desempenho abaixo do esperado em consequência do algoritmo de decisão baseado em um esquema de voto majoritário.

Na segunda implementação, conforme a nossa proposta de pesquisa, utilizamos a técnica da energia das silhuetas do caminhar associada com a proporção de variâncias para permitir, através da imagem contendo os índices de energia das silhuetas, a fusão das características com as imagens faciais. Neste caso, avaliou-se e constatou-se a viabilidade da proposta, onde pudemos confirmar um bom desempenho do classificador RBF na fusão das características de ambos indicadores biométricos para uma determinada faixa de coeficientes característicos.

# Capítulo 6

## **As Bases de Dados do CEFETES e a Fusão no Nível das Características**

A obtenção de bases de dados próprias contendo seqüências de vídeo da forma de caminhar e de imagens faciais exige equipamentos e cenários adequados e, principalmente, por envolver imagens de pessoas, não é uma tarefa nada fácil. Nosso propósito inicial era de obter bases de dados que pudessem representar, para a mesma seqüência de vídeo do indivíduo caminhando, um quadro da seqüência e uma imagem facial capturados de forma simultânea. Desta forma, se considerarmos uma pessoa cujo ciclo do caminhar contém 40 quadros capturados à taxa de 30 quadros/seg, teríamos associadas 40 imagens faciais com as silhuetas do caminhar a serem extraídas.

Devido às limitações apresentadas pela interface de aquisição, que diminuía a taxa de captura para 15 quadros/seg com a captura simultânea, e à necessidade de utilização de algoritmos para localização automática das faces em movimento, implicando em ajustes automáticos dos ângulos e alturas da câmara e, em consequência, também do tempo reduzido, optamos por construir bases de dados com as imagens faciais obtidas no final da seqüência do caminhar e com o indivíduo parado diante da câmara. Sendo assim, apesar de tantos outros contratempos e limitações, propusemos e obtivemos, com a colaboração de um bom número de voluntários formados por alunos e servidores do CEFET-ES, quatro bases de dados cujas capturas foram feitas de duas a duas para faces e forma de caminhar, em épocas distintas para cada indivíduo e separadas em intervalos de tempo de alguns dias.

## 6.1 As etapas de aquisição e as características das bases de dados

Inicialmente, foi necessário constituir um número razoável de voluntários. Dessa forma, 58 pessoas participaram na formação das duas primeiras bases de dados de seqüências de vídeo do caminhar e de imagens faciais. Então, foi montado no respectivo laboratório o cenário formado por duas câmaras de vídeo fixadas ao longo de um corredor que se encerra em uma porta de acesso restrito, de acordo com a sugestão apresentada pela Figura 35, e iniciou-se a captura das seqüências de vídeo necessárias ao levantamento dos dados das duas primeiras bases de dados, uma para as seqüências de vídeo do caminhar e outra para as respectivas imagens faciais.

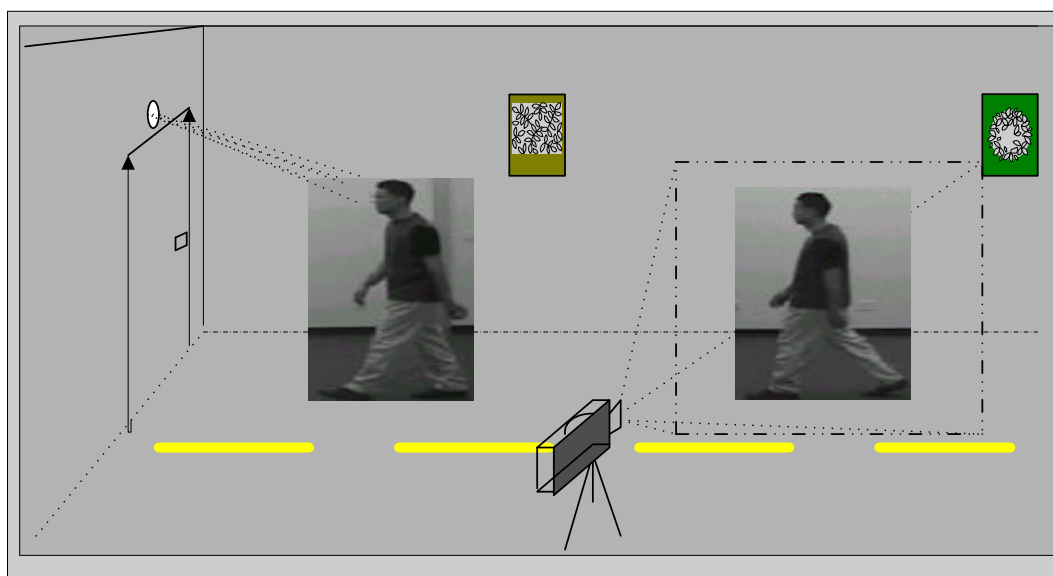


Figura 35 – Cenário proposto para a fusão face-forma de caminhar

Tal levantamento foi feito de acordo com a disponibilidade de cada participante, a qualquer hora do dia, e com a iluminação proporcionada por lâmpadas fluorescentes. Além disso, a captura das seqüências da forma de caminhar e das imagens faciais ocorreu ao longo de alguns dias e sem a preocupação rigorosa com pequenas alterações das posições e alturas relativas das câmaras filmadoras, uma vez que o cenário era desmontado após um conjunto diário de aquisições.

O procedimento para a captura das seqüências de vídeo da forma de caminhar e das imagens faciais para cada pessoa era tal que, após a obtenção das seis seqüências de vídeo do caminhar, estabelecia-se a captura das imagens faciais.

Alem disso, simultaneamente com a obtenção da base de dados, tão logo um determinado número de características era obtido, iniciava-se o pré-processamento das mesmas.

Por fim, foram levantados os dados das seqüências do caminhar e imagens faciais de mais duas bases de dados, agora com a participação de 54 voluntários, que haviam participado também das duas primeiras bases de dados.

Em todas as etapas, as seqüências de vídeo do caminhar e as imagens faciais correspondentes para cada indivíduo, conforme os exemplares apresentados na Figura 36, foram obtidas a uma taxa de captura de 30 quadros por segundo, de onde foram extraídos os quadros em tons de cinza com uma resolução de 320 x 240 pixels.



Figura 36 – Exemplares das bases de dados do CEFET-ES

As silhuetas do caminhar, em preto-e-branco, foram formadas por quadros correspondentes a um ciclo do caminhar e, em seguida, foram normalizadas e reduzidas para uma resolução de 141 x 231 pixels. As imagens faciais em tons de cinza, por sua vez, foram reduzidas e normalizadas para uma resolução de 141 x 186 pixels.

Os algoritmos utilizados na classificação das características obtidas nas duas primeiras bases de dados não sofreram nenhuma alteração em relação aos procedimentos do capítulo anterior, sendo utilizados os mesmos “scripts” implementados no Matlab®. Neste caso, foram obtidos os vetores de características

básicos contendo 32571 amostras para cada imagem ESC correspondente a uma seqüência do caminhar e 26226 amostras para o vetor de característica da imagem facial correspondente àquela seqüência.

Tendo-se em conta que o desempenho de um sistema de reconhecimento pessoal, como apresentado no capítulo 2, pode ser analisado considerando-o como um sistema de identificação ou como um sistema de verificação, iremos apresentar nas próximas seções os resultados obtidos com a classificação dos vetores de características originados nas bases de dados do CEFET-ES e analisar as diversas curvas contendo as taxas de acertos, de falsas aceitações e falsas rejeições.

## **6.2 Resultados considerando o sistema no modo de identificação**

Neste estudo, para a obtenção dos resultados do sistema no modo de identificação, se utilizou das bases de dados integrais, sem nenhuma informação de identidade. Neste caso, para as duas primeiras bases de dados, os resultados obtidos com o classificador RBF são tais que comparam uma seqüência do caminhar, uma face ou a fusão de ambas (seqüência do caminhar-face) com todas as demais 347 seqüências restantes correspondentes às 58 classes de indivíduos, em um processo de validação cruzada.

### **6.2.1 Resultados obtidos com os dados da 1ª fase**

A Figura 37 apresenta as taxas de acertos para a situação anterior. Ela considera uma taxa de eliminação de coeficientes menos discriminatórios sobre os vetores de características da fusão face-forma de caminhar dentro da faixa de 0,10 a 0,25 dos valores de  $r_m$  normalizados. Esta faixa de valores corresponde a uma variação na dimensão dos vetores que varia desde 850 até 5550 amostras. Neste caso, novamente aqui, devido à não linearidade de  $r_m$ , observamos que o número de coeficientes dos vetores de características não se altera de forma linear.

Através das curvas da Figura 37, observamos que as taxas de acertos para as imagens faciais em toda a faixa considerada é relativamente baixa. Isso deve ter ocorrido pelo fato das imagens terem sofrido o corte e um único processo de normalização baseado em escala. Observamos, também, que o melhor desempenho do classificador, para as três curvas das Figuras 37, ocorre quando a eliminação de



coeficientes menos discriminatórios dos vetores de características é tal que o número de amostras está dentro de uma faixa de 2000 a 4500.

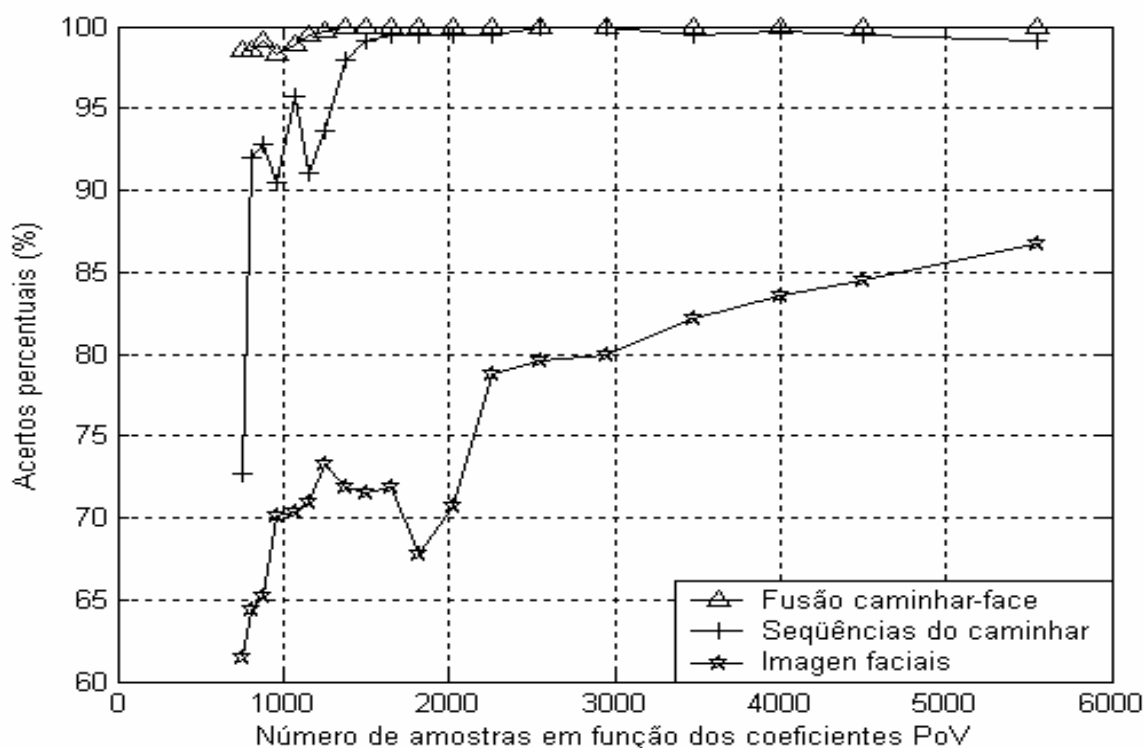


Figura 37 – Resultados de classificação com exemplares de dados da 1ª fase

Observa-se, ainda, que a curva com as taxas de acertos para a fusão das duas características apresenta um ótimo desempenho quando comparada às curvas das características individuais. Em especial, observamos que, para vetores de características contendo por volta de 1200 coeficientes, quando a taxa de acertos é baixa para as imagens faciais e a forma de caminhar, a fusão das duas foi capaz de gerar taxas de acertos da ordem de 100%. Isto demonstra a eficácia da estratégia de fusão no nível de características da nossa proposta.

### 6.2.2 Resultados obtidos com os dados da 2ª fase em relação à 1ª fase

Os algoritmos e “scripts” necessários à obtenção dos resultados de classificação das bases de dados obtidas na segunda fase, contendo 54 indivíduos, quando comparadas para fins de classificação com as bases de dados da fase anterior, contendo 58 indivíduos, exigiram algumas alterações tanto para as características da forma de caminhar quanto para as das imagens faciais.

Neste caso, após a aplicação da redução de dimensionalidade através da técnica PoV nos dados da 1ª etapa, se obteve um vetor de índices contendo a ordem dos coeficientes de eliminação PoV do vetor de características. Tal procedimento foi necessário para se obter a redução equivalente na 2ª etapa de forma que os coeficientes, que refletem as características correspondentes nas colunas da matriz, que contêm os vetores de características na 1ª etapa, sejam aqui também considerados. Sendo assim, o vetor ordenado de índices PoV foi salvo e armazenado na 1ª fase para permitir que ambas as bases de dados tenham seus coeficientes projetados em um mesmo sub-espço.

Considerando o sistema operando no modo de identificação, os resultados para a menor distância entre os escores, retornados pelo classificador em relação ao padrão de teste, são apresentados na Figura 38. Eles são tais que comparam um vetor de coeficientes PoV da sequência do caminhar, de uma face ou da fusão de ambas. Para tanto, o algoritmo retira um dos 324 vetores de características da 2ª fase e compara-o com todos os demais 348 vetores das seqüências correspondentes às 58 classes de indivíduos obtidos na 1ª fase.

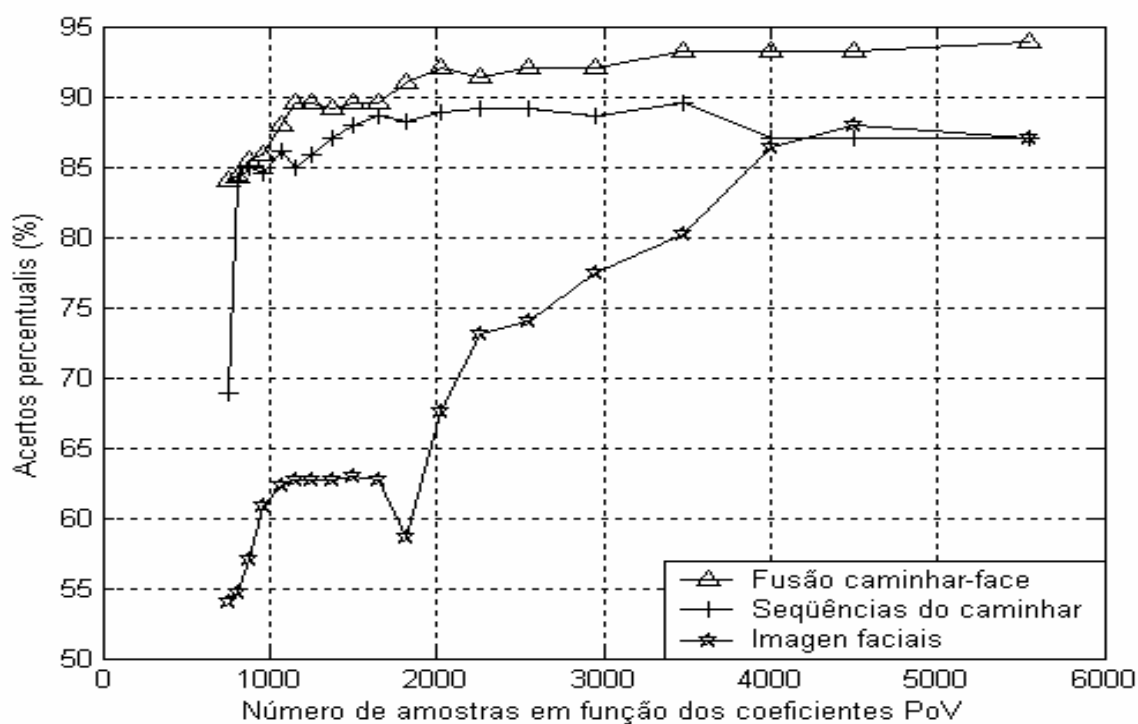


Figura 38 – Resultados de classificação com exemplares da 2ª fase x 1ª fase

A obtenção das curvas através do classificador neural RBF implementado no Matlab® levou em consideração a função de similaridade que avalia as classes e retorna os escores obtidos entre os padrões de teste e de treinamento. Neste caso, foi tomada uma faixa de escores que permitiu estabelecer dois limiares de corte (baixo = 0,0 e alto = 1,50), de tal forma que os escores retornados fora desta faixa seriam atribuídos a uma classe de indivíduos genuínos não reconhecidos.

Ao comparar as curvas das Figuras 37 e 38, apesar das taxas de acertos inferiores para todas as situações expressadas pelas curvas da 2ª etapa, observamos similaridades entre elas. Em ambas, as taxas de acertos para as imagens faciais foram relativamente baixas, sendo hipóteses possíveis para este fato, o pouco pré-processamento de normalização das imagens faciais e a baixa capacidade de discriminação oferecida pelo grande número de coeficientes dos vetores de características para o classificador RBF.

Observamos, ainda, que as taxas de acertos para a forma de caminhar foram ligeiramente superiores às das imagens faciais. Entretanto, na fusão das características em ambas as etapas, as taxas de acertos foram superiores a ambas, confirmando um melhor desempenho da nossa proposta.

### **6.3 Análise das taxas de acertos em função dos coeficientes PoV**

A análise de desempenho do sistema em função dos coeficientes PoV, no que diz respeito à aceitação ou rejeição de indivíduos genuínos, apesar do pequeno número de indivíduos nas bases de dados, também foi realizada. Para tais testes foi estabelecida uma faixa de escores compreendida entre dois limiares de corte baixo e alto (0,0 e 1,4, respectivamente), de tal forma que os escores retornados fora desta faixa seriam atribuídos a uma classe de indivíduos genuínos não reconhecidos. A finalidade dos testes não é a de obter taxas excelentes na classificação, visto que o objetivo principal do trabalho é comprovar a nossa hipótese de um melhor desempenho de classificação em função da proposta de fusão.

Desta forma, considerando que dispomos em nossas bases de dados das verdadeiras identidades dos indivíduos, então podemos avaliar os resultados do sistema supondo sua operação no modo de identificação ou de verificação. Para o

primeiro caso, a Figura 39 apresenta as curvas contendo as taxas de acerto do classificador RBF e a taxa de identidades genuínas rejeitadas, ambas em função do número de coeficientes por vetor de características. Neste caso, a análise é feita só para a fusão das características de seqüências do caminhar e imagens faciais relativas à comparação entre os dados da 1ª fase.

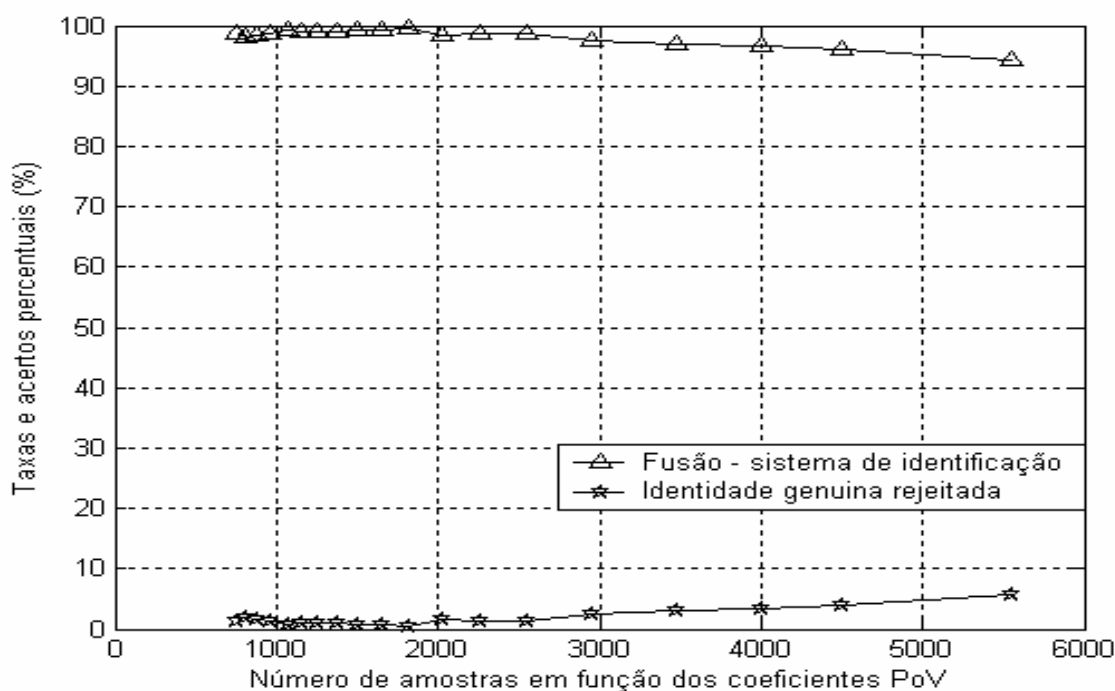


Figura 39 – Sistema de identificação: curvas para a fusão com as taxas referentes à 1ª fase

A seguir, considerando o comportamento do sistema como de verificação e as mesmas condições para os limiares de corte nos escores retornados pelo classificador RBF para a fusão das características, a Figura 40 apresenta agora os resultados de classificação sintetizando a comparação dos 324 vetores de características de teste das bases de dados da 2ª fase contra os 348 vetores de treinamento da 1ª fase, respectivamente.

O objetivo principal é o de avaliar alguns indícios do comportamento do sistema de reconhecimento em função do número de coeficientes PoV e, para tanto, neste gráfico, são também calculadas e apresentadas as taxas de acertos do classificador RBF, as taxas de identidades falsamente rejeitadas, bem como as taxas de identidades aceitas de indivíduos genuínos, porém com as classes trocadas.

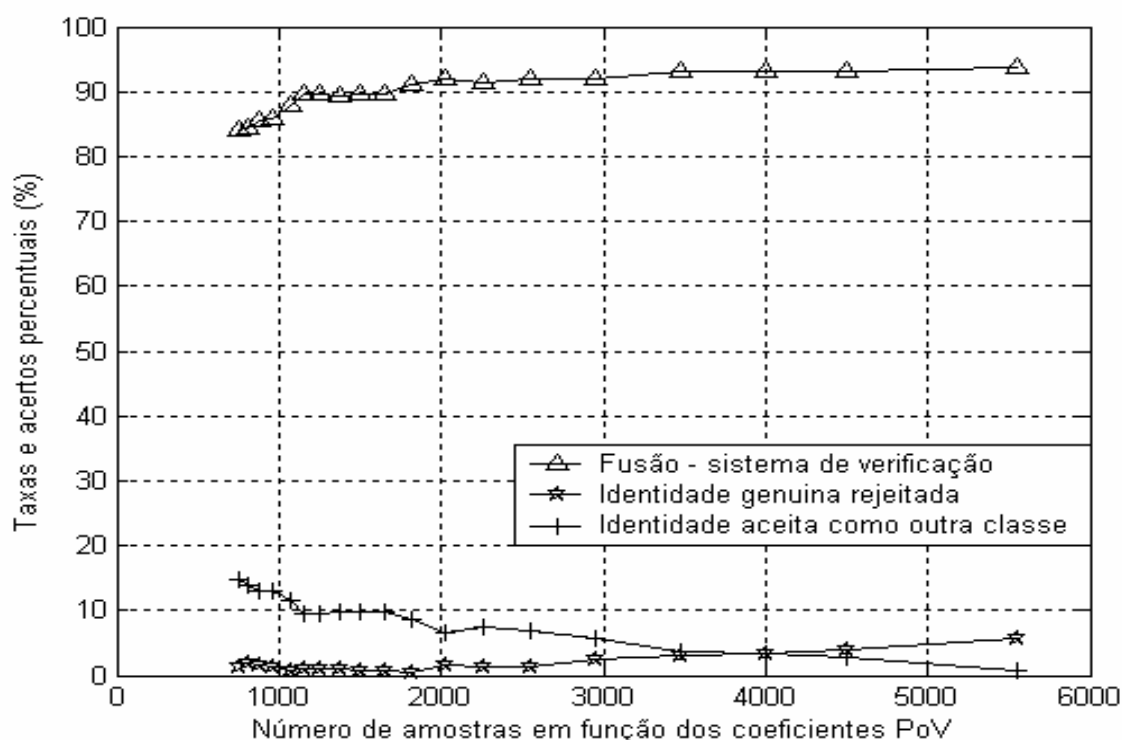


Figura 40 – Sistema de verificação: curvas com as taxas de acertos, identidade rejeitadas e identidades aceitas como de outra classe

Observamos, na Figura 40, que as duas curvas de rejeição de identidades genuínas e aceitação trocada de identidades genuínas têm um comportamento tal que, enquanto a primeira cresce, a outra decresce com o aumento do número de coeficientes PoV interceptando-se em um ponto no qual temos aproximadamente 4000 coeficientes PoV na representação dos vetores de características, o que nos indica um ponto a ser usado como referência para as próximas análises.

#### 6.4 Análise das taxas de falsa aceitação e falsa rejeição

Para esse estudo, o sistema deve operar no modo de verificação e a análise das taxas falsa aceitação (FAR) e falsa rejeição (FRR) foi feita de acordo com o protocolo utilizado em BioID (2007), no que diz respeito à escolha dos limiares de escores retornados pela função de similaridade do classificador RBF (FRISCHHOLZ; DIECKMANN, 2000).

As bases de dados contendo as características da fusão biométrica foram particionadas em três subconjuntos, onde o primeiro, contendo 30 indivíduos, foi extraído da base de dados obtida na 1ª fase e utilizado no treinamento do

classificador RBF; o segundo e o terceiro conjuntos, contendo 30 e 24 indivíduos, respectivamente, foram extraídos das bases de dados obtida na 2ª fase e utilizados para a obtenção das taxas de falsa aceitação (FAR – *False Acceptance Rate*) e de falsa rejeição (FRR – *False Rejection Rate*). A partir disso, para a fusão das características, avaliamos o comportamento do sistema em três situações, uma inferior contendo 3000 coeficientes PoV, outra superior com 5000 e, por último, a análise considerando 4000 coeficientes PoV por vetor.

A Figura 41 apresenta, para vetores de características contendo 3000 coeficientes PoV, as respectivas curvas de FAR e FRR e o ponto em que as duas taxas se igualam denotado por EER (*Equal Error Rate*).

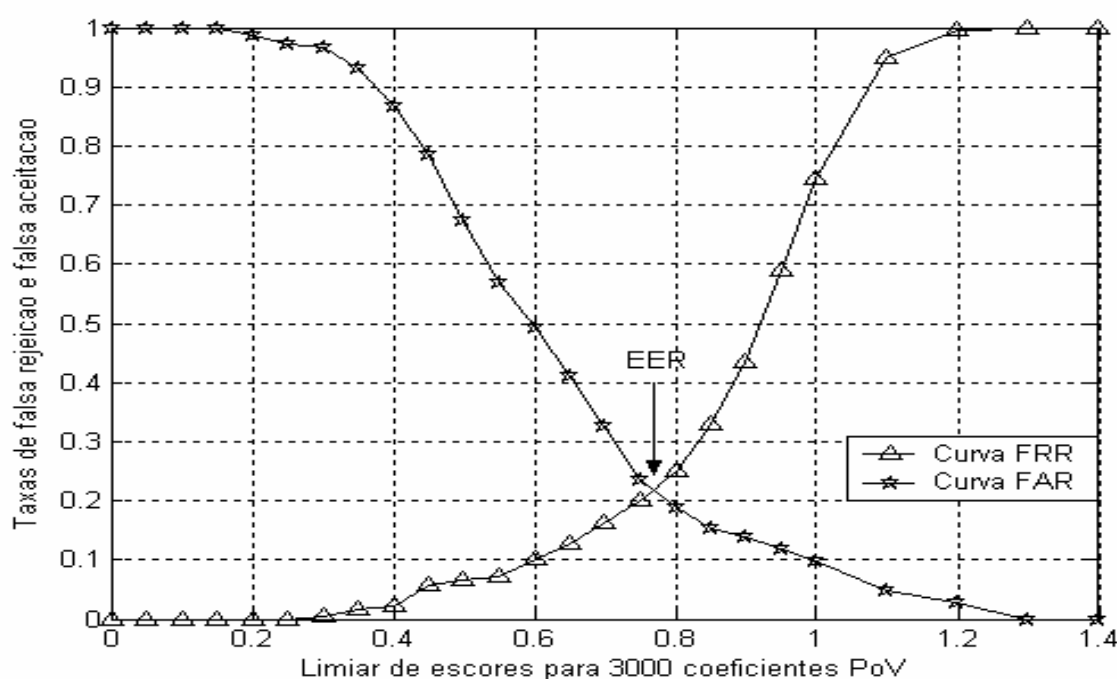


Figura 41 – Curvas FAR e FRR: 3000 coeficientes na fusão face-forma de caminhar

Neste ponto, é importante ressaltar que, em geral, para aplicações comerciais, procura-se ajustar o ponto EER para a operação do sistema de verificação (JAIN; ROSS, PRABHAKAR, 2004).

A Figura 42, por sua vez, apresenta as curvas de falsa rejeição (FRR) e falsa aceitação (FAR) para os vetores de características contendo 5000 coeficientes PoV.

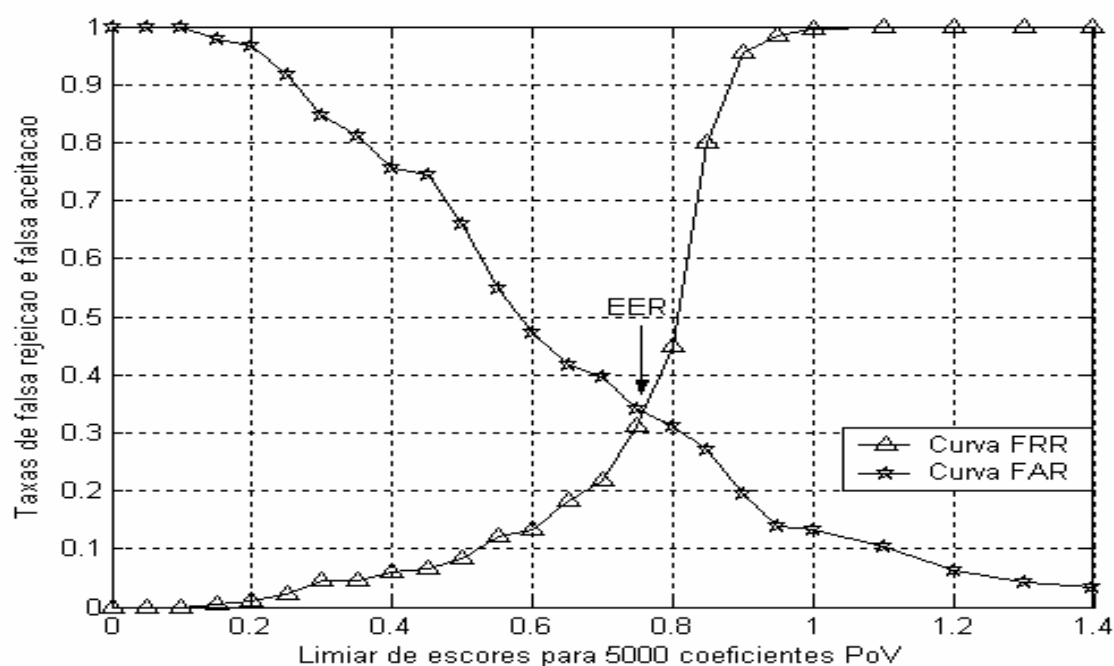


Figura 42 – Curvas FRR e FAR: fusão face-forma de caminhar considerando 5000 coeficientes PoV

A Figura 43 apresenta as curvas FRR e FAR para vetores de características contendo 4000 coeficientes PoV da fusão imagens faciais-sequências do caminhar.

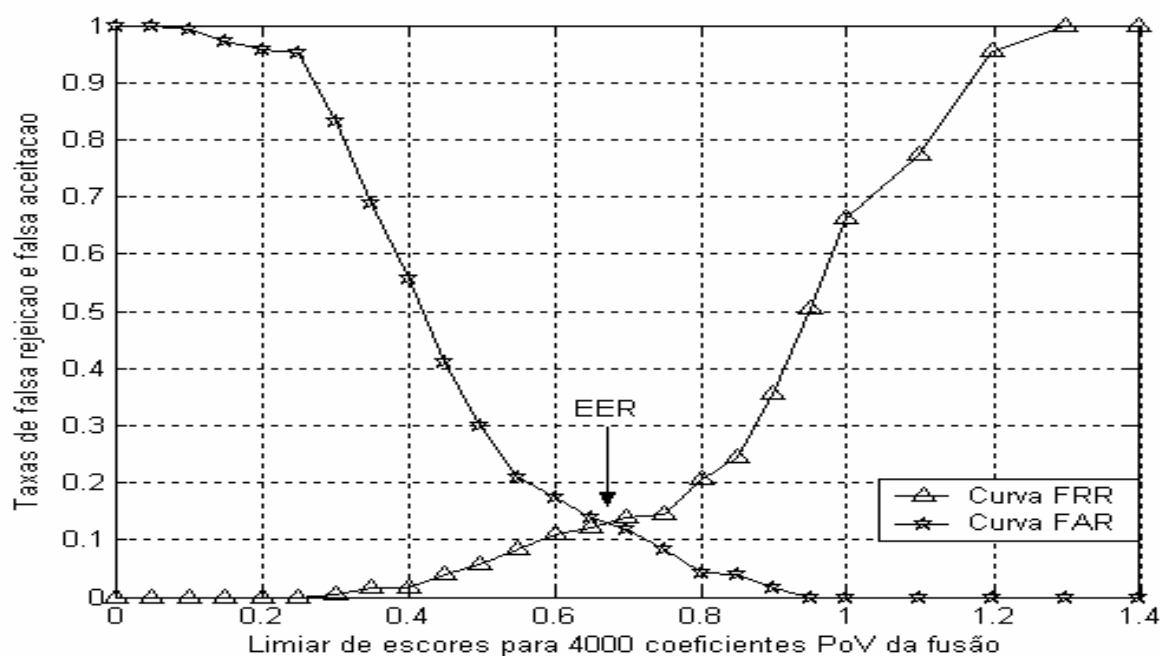


Figura 43 – Curvas FRR e FAR: fusão face-forma de caminhar considerando 4000 coeficientes PoV

Em função dos bons resultados apresentados com os vetores de características contendo 4000 amostras para o sistema operando no modo de verificação, fizemos

um estudo do seu comportamento também para as imagens faciais e as seqüências de vídeo do caminhar.

A Figura 44 apresenta as curvas de FRR e FAR para vetores de características faciais contendo 4000 coeficientes PoV.

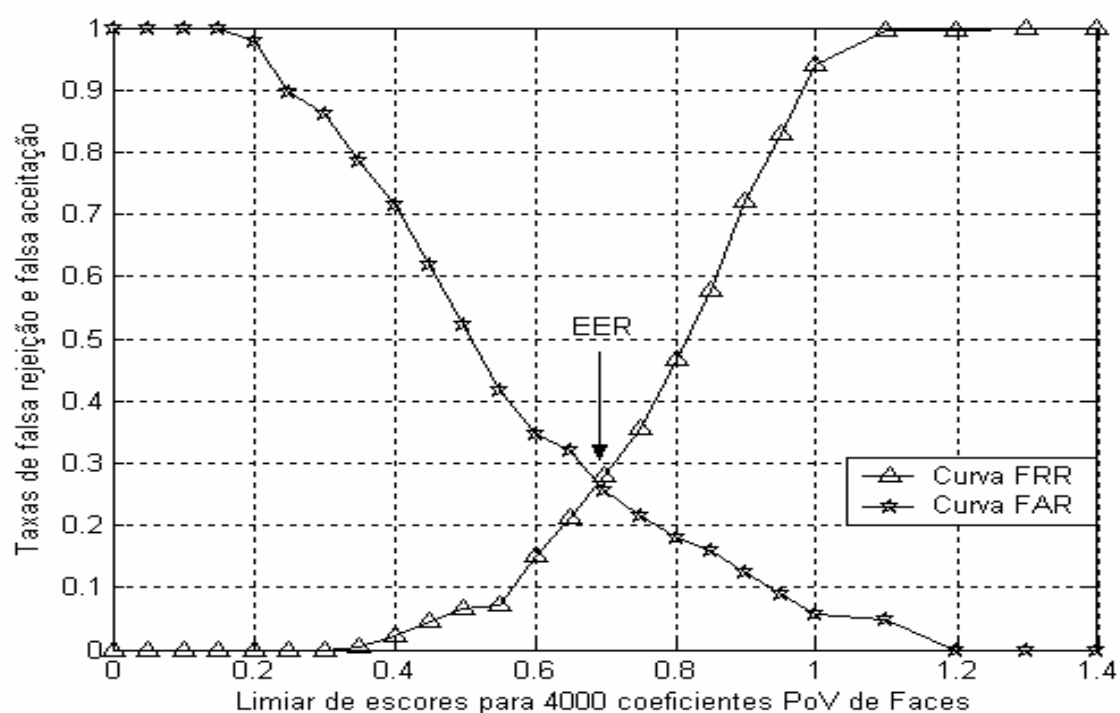


Figura 44 – Curvas FRR e FAR: imagens faciais considerando 4000 coeficientes PoV

Observamos, neste caso, como era esperado diante do baixo desempenho de classificação obtido anteriormente para as imagens faciais, que a taxa de erros iguais é da ordem de 0,27, a qual é bastante elevada.

A Figura 45 apresenta as curvas de FRR e FAR para vetores de características de seqüências do caminhar contendo 4000 coeficientes PoV.



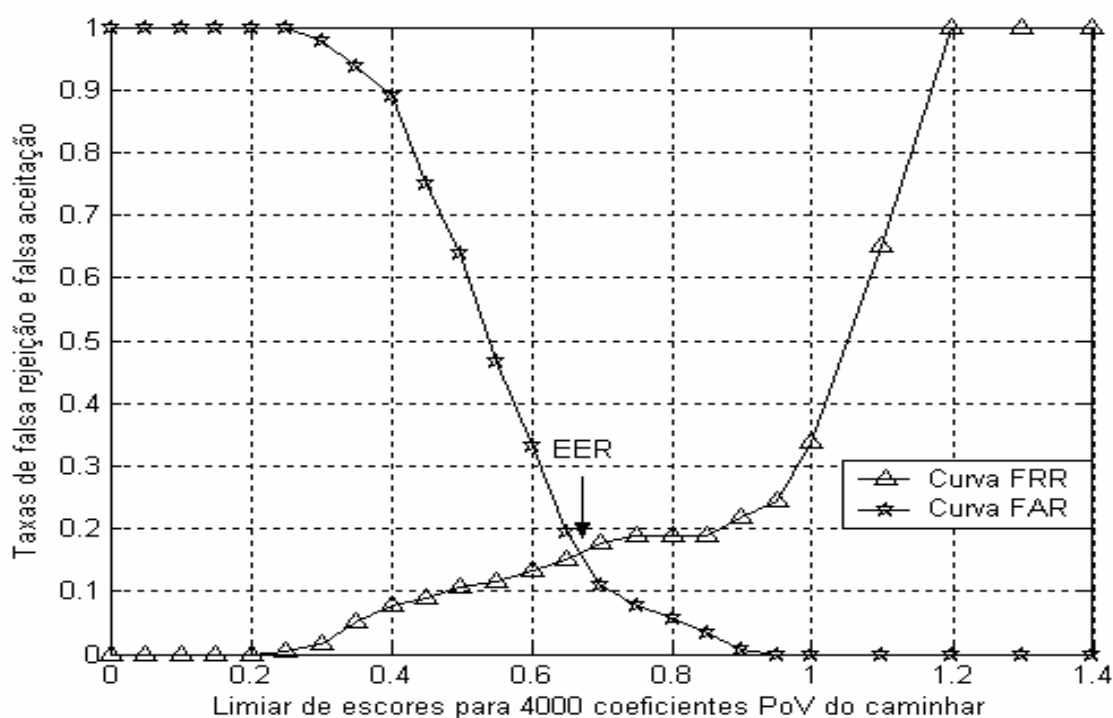


Figura 45 – Curvas FRR e FAR: seqüências do caminhar considerando 4000 coeficientes PoV

Aqui, observamos, através da taxa de erros iguais (*EER – Equal Error Rate*) de aproximadamente 0,15, um melhor desempenho de classificação para as seqüências de vídeo da forma de caminhar em relação às imagens faciais.

Ao compararmos as Figuras 43, 44 e 45 vemos claramente que o melhor desempenho de classificação ocorre com as curvas da Figura 43, na qual são apresentados os resultados da fusão no nível das características dos vetores das imagens faciais e das seqüências de vídeo do caminhar. Neste caso, o ponto em que as curvas se cruzam com erros iguais (*EER – Equal Error Rate*) e atingem ambas aproximadamente a taxa de 0,11, se dá para limiares de escores retornados pela função de similaridade do classificador RBF superiores a 0,7.

## 6.5 Considerações finais

Neste capítulo, apresentamos a base de dados obtidas no CEFET-ES e os resultados obtidos com a fusão entre as características faciais e as seqüências da forma de caminhar.

Ao comparar as curvas das Figuras 35, 37 e 38 que apresentam os resultados para as bases de dados de domínio público OLIVETTI/GATECH e proprietária CEFET-ES, observamos que, apesar das diferentes taxas de acertos para todas as situações expressas nas curvas, há similaridades entre elas. Neste caso, em todos os gráficos, as taxas de acertos para as imagens faciais foram relativamente baixas, e as taxas de acertos para a forma de caminhar foram superiores às das imagens faciais. Entretanto, confirmando a nossa hipótese inicial de que haveria um melhor desempenho quando se implementasse a fusão no nível das características, em todas as etapas as taxas de acertos face-forma de caminhar foram superiores às demais.

Os resultados obtidos com as bases de dados do CEFET-ES permitiram avaliar o desempenho e alguns aspectos de robustez do sistema de reconhecimento sob os dois pontos de vista biométricos, isto é, considerando-o como de identificação e, em seguida, como de verificação. Assim, para ambas as considerações foi possível a análise das taxas de acertos, de identidades falsamente aceitas e rejeitadas em função do número de coeficientes PoV, como também em função dos limiares de corte estabelecidos pelos escores retornados pela função de similaridade do classificador RBF.

Observamos o comportamento do sistema em função dos coeficientes PoV, ao considerá-lo no modo de verificação e efetuando a classificação das bases de dados da 2ª fase contra as da 1ª fase (54 x 58 classes). Neste caso, as taxas com erros iguais nas curvas de identidades falsamente aceitas e identidades falsamente rejeitadas ocorreram para vetores de características contendo 4000 coeficientes.

De uma forma geral, apesar das limitações na captura dos dados e do número reduzido de classes, os resultados obtidos através das duas implementações propostas, conforme apresentam as Figuras 45, 44 e 43, respectivamente, as curvas com as taxas FAR, FRR e EER para os três sistemas constituídos de vetores de características da forma de caminhar, de imagens faciais e da fusão de ambas, que a arquitetura apresentada é factível e que a nossa tese da fusão dos dois indicadores biométricos e comportamentais no nível das características é válida, pode ser efetuada de forma bastante simples e apresenta um melhor desempenho.

Para finalizar, ao compararmos a metodologia proposta com os resultados apresentados nos trabalhos de Kale et al, (2002), Shakhnarovich, Lee e Darrell (2001) e Shakhnarovich e Darrell (2002), cuja fusão foi implementada em outros níveis para os dois sistemas biométricos constituídos pela forma de caminhar e imagens faciais, verificamos que, embora eles não apresentassem uma análise das taxas FRR e FAR, nossos resultados são iguais ou superiores àqueles, o que nos garante a viabilidade da proposta aqui apresentada.

# Capítulo 7

## Conclusão

Neste trabalho de pesquisa, em uma primeira etapa propusemos algumas abordagens que permitiram avaliar o desempenho das características biométricas e comportamentais da forma de caminhar na implementação de sistemas de reconhecimento pessoal. O nosso objetivo principal, além de avaliar o desempenho das diversas arquiteturas de classificadores em função das técnicas de extração e seleção de características, era, devido ao ineditismo do tema, propor o uso de técnicas que permitissem a fusão dos dois indicadores biométricos face e forma de caminhar no nível das características. Sendo assim, dado o nosso interesse em trabalhar e avaliar o desempenho de sistemas biométricos que permitem a identificação à distância e com pouca ou nenhuma colaboração do indivíduo, duas arquiteturas capazes de efetuarem a fusão das características das imagens faciais e das silhuetas de seqüências da forma de caminhar foram propostas e avaliadas.

Em todas as implementações, o desempenho da fusão no nível de características foi superior àquele em que consideramos somente as características da forma de caminhar ou das imagens faciais, ou ainda, quando a fusão ocorreu no nível dos classificadores. Adicionalmente, para uma determinada faixa de coeficientes dos vetores de características, a fusão ocorrendo no nível das características ofereceu um desempenho bem superior para ambas as bases de dados consideradas no estudo, a de domínio público OLIVETTI/GATECH e a proprietária CEFET-ES.

Os resultados obtidos com as bases de dados do CEFET-ES permitiram avaliar o desempenho e também alguns aspectos de robustez do sistema de reconhecimento sob os dois pontos de vista biométricos, isto é, considerando o sistema operando no modo de identificação e, em seguida, no modo de verificação. Neste caso, para ambas as considerações, foi possível a análise das taxas de acertos, de identidades

falsamente aceitas e rejeitadas em função do número de coeficientes PoV, o que permitiu, em 4000 coeficientes, encontrar o ponto de taxas de erros iguais. Em seguida, foram analisadas as taxas de falsa aceitação (FAR) e falsa rejeição (FRR) em função dos limiares de corte estabelecidos nos escores retornados pela função de similaridade do classificador RBF. Aqui, observamos que os vetores de características com 4000 coeficientes apresentaram, para a fusão dos dois indicadores, a menor taxa com erros iguais (EER) nas curvas de falsa aceitação (FAR) e falsa rejeição (FRR).

### **7.1 Sumário de contribuições**

Entre nossas contribuições destacamos, em especial, o bom desempenho das arquiteturas e algoritmos propostos para a fusão face-forma de caminhar:

- a primeira, no nível de decisão, em que o classificador da forma de caminhar funciona como um filtro emitindo os escores de possíveis identidades que têm a vantagem de possibilitar uma pré-identificação pessoal à distância, de forma não-intrusiva e sem a colaboração do indivíduo. Entretanto, ela não apresentou um desempenho aceitável em consequência da decisão final que ocorreu entre dois escores e por voto majoritário;
- a segunda, no nível de características, onde utilizamos a técnica de extração de características da Energia das Silhuetas do Caminhar em composição com a Proporção de Variâncias para permitirem, através da imagem com os índices de energia das silhuetas, a redução de dimensionalidade e a fusão das características com as imagens faciais. Em função da baixa complexidade do algoritmo e do seu baixo custo computacional, apesar da necessidade de se dispor simultaneamente dos vetores de características de ambos os domínios biométricos, tal proposta apresentou resultados inéditos e animadores.

Destacamos também, as técnicas, arquiteturas e algoritmos desenvolvidos ou adaptados para os sistemas de reconhecimento pessoal com base na forma de caminhar. Além disso, dentro do escopo deste trabalho, destacamos ainda:

- a adaptação e aplicação de técnicas estatísticas na obtenção do fundo do cenário a partir de quadros da própria seqüência de vídeo do caminante;
- a adequação da técnica da energia das silhuetas do caminhar, que permite reduzir o custo computacional na fase de pré-processamento e ainda representar, em uma única imagem, aspectos espaço-temporais das seqüências de vídeo da forma de caminhar;
- a adequação, aplicação e avaliação, entre outras, das técnicas de extração e seleção de características baseadas em Proporção de Variâncias, Transformadas Wavelets e Análise de Componentes Independentes;
- a adequação da técnica de seleção baseada na proporção de variâncias para reduzir a dimensionalidade e para normalizar os vetores de características no processo de fusão bi-modal, que, neste caso, faz uma ordenação dos coeficientes com todas as amostras dos vetores de características e, em seguida, em função de um limiar de corte, exclui os coeficientes de menor poder discriminatório;
- a adequação e aplicação de classificadores baseados em Máquinas de Vetores Suporte (SVM) e em Redes Neurais com retro-propagação e com Funções de Base Radiais (RBF).

## **7.2 Pesquisas futuras**

Devido a possíveis correlações entre as características das imagens faciais e as seqüências de vídeo da forma de caminhar, quando obtidas simultaneamente para uma mesma pessoa, no escopo deste trabalho, planejamos ainda desenvolver algoritmos de busca para obter simultaneamente, nas bases de dados bi-modais, seqüências de vídeo sincronizadas das imagens faciais e da forma de caminhar do indivíduo caminhando diante das câmaras e, em seguida, reavaliar o desempenho dos sistemas de reconhecimento pessoal propostos ao efetuar-se a fusão no nível das características dos dois indicadores biométricos face-forma de caminhar.

Dado a flexibilidade da arquitetura proposta e os bons resultados alcançados com dados de domínio público para as imagens faciais termográficas, conforme Salomão,

Almeida e Salles (2006a), tão logo seja possível, pretendemos obter bases de dados no espectro infravermelho e reavaliar todas as arquiteturas e algoritmos da nossa proposta.

Em futuros trabalhos, pretendemos efetuar novos estudos da utilização da nossa proposta ao considerar-se outros indicadores biométricos, entre os quais podemos citar a fusão das imagens faciais, do movimento labial e da voz.

## REFERÊNCIAS

- ABE, S.; INOUE, T. Fuzzy Support Vector Machines for Multiclass Problems. *European Symposium on Artificial Neural Networks*. Bruges – Belgium, 2002, p. 113-118.
- BARCLAY, C.D.; CUTTING, J.E.; KOZLOWSKI, L.T. Temporal and Spatial Factors in Gait Perception that Influence Gender Recognition. *Perception and Psychophysics*, n. 23, 1978, p.145–152.
- BARTLETT, M. S.; MOVELLAN, J. R.; SEJNOWSKI, T. J. Face Recognition by Independent Component Analysis. *IEEE Transactions on Neural Networks*, v. 13, n. 6, 2002, p. 1450-1464
- BELL, A.; SEJNOWSKI, T.J. An Information-Maximization Approach to Blind Separation and Blind Deconvolution. *Neural Computation*, v. 7, 1995, p. 1129-1159.
- BENABDELKADER, C.; DAVIS, L.; CUTLER, R. Motion Based Recognition of People in Eigengait Space. *In Proceedings IEEE International Conference on Automatic Face and Gesture Recognition*, 2002, p. 267-272.
- BIOID HumanScan. Disponível em: [http://www.bioid.com/sdk/docs/About\\_EER.htm](http://www.bioid.com/sdk/docs/About_EER.htm). Acesso em: 17 jul. 2007.
- BISHOP C. M. *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.
- BIUK, Z.; LONCARIC, S. Face Recognition from Multi-pose Image Sequence. *In Proc. of 2nd International Symposium on Image and Signal Processing and Analysis*. Pula, Croatia, 2001, p. 319-324.
- BOBICK, A.F.; DAVIS, J.W. The Recognition of Human Movement using Temporal Templates. *IEEE Trans. Pattern Analysis and Machine Intelligence*, v. 23, n. 3, 2001, p. 257-267.
- BOULGOURIS, N. V.; HATZINAKOS, D.; PLATANIOTIS, K. N. Gait Recognition: a Challenging Signal Processing Technology for Biometric Identification. *IEEE Signal Processing Magazine*, v. 22, n. 6, 2005, p. 78-90.
- BURGESS, C.J.C. A tutorial on Support Vector Machines for Pattern Recognition. *Data Mining Knowledge Discovery*, Kluwer Academic Publishers, Boston, n.2, 1998, p. 121 –167.
- BURRUS, C. S.; GOPINATH, R. A.; GUO, H., Introduction to Wavelets and Wavelet Transform – a Primer, *Prentice Hall*, New Jersey, 1998.
- CANU, S.; GRANDVALET, Y.; RAKOTOMAMONJY, A. SVM and Kernel Methods Matlab Toolbox. *Perception Systèmes et Information*, INSA, Rouen, France, 2003.



- CASIA Gait Database. Disponível em: <<http://www.inobiometrics.com>>. Acesso em: 20 mar. 2005.
- CAWLEY, G. C.; TALBOT, N. L. C. Efficient Leave-One-Out Cross-Validation of Kernel Fisher Discriminant Classifiers. School of Information Systems, University of East Anglia, Norwich, United Kingdom, NR4 7TJ, Preprint submitted to Elsevier Science, 2003.
- CHELLAPPA, R.; ROY-CHOWDHURY, A. K.; ZHOU S.K. Human Identification Using Gait and Face. Disponível em: <<http://citeseer.ist.psu.edu/715072.html>>. Acesso em: 20 mar.2005.
- CHOUDHURY, T.; CLARKSON, B.; JEBARA, T.; PENTLAND, A. Multimodal Person Recognition using Unconstrained Audio and Video. In *Proceedings, International Conference on Audio- and Video-Based Person Authentication*, Washington-DC, 1999, p. 176–181.
- COLLINS, R.; GROSS, R.; SHI, J. Silhouette-based Human Identification from Body Shape and Gait. In: *Proc.5th IEEE Face Gesture Recognition*, 2002, p. 366 - 371.
- CUNADO, D.; NASH, J.M.; NIXON, M.S.; CARTER, J. N. Gait Extraction and Description by Evidence-Gathering. *Proc. Of the International Conference on Audio and Video Based Biometric Person Authentication*, 1999, p. 43–48.
- CUNADO, D.; NIXON, M. S.; CARTER, J. N. Automatic Gait Recognition via Model-Based Evidence Gathering. In: *Proc. AutoID99: IEEE Workshop on Automated ID Technol.*, 1999a, p. 27 - 30.
- CUTTING, J.; KOZLOWSKI, L. Recognizing Friends by their Walk: Gait Perception without Familiarity Cues. *Bulletin of the Psychonomic Society*, n. 9, 1977, p. 353–356.
- CUTTING, J.E.; PROFFITT, D.R. Gait Perception as an Example of how we may Perceive Events. *Plenum Press*, London, 1981.
- DAVIS, J. W.; TAYLOR, E. R. Analysis and Recognition of Walking Movements. *International Conference on Pattern Recognition*, Quebec, Canada, 2002, p. 315-318.
- DITTRICH, W.H. Action Categories and the Perception of Biological Motion. *Perception*, n. 22, 1993, 15–22.
- DUDA R. O.; HART P. E.; STORK D. G., Pattern Classification. 2nd Edition, *Wiley-Interscience*, 2000.
- EKENEL, H. K.; SANKUR, B. Feature Selection in the Independent Component Subspace for Face Recognition. *Pattern Recognition Letters*, n.25, 2004, p.1377 –1388.
- FLIEGE, N.J. *Multirate Digital Signal Processing*. New York: John Wiley & Sons Ltda, 2000.

- FOSTER, J. P., NIXON M. S., PRÜGEL-BENNETT, A. Automatic Gait Recognition Using Area-Based Metrics. *Pattern Recognition Letters*, n. 24, 2003, p. 2489-2497.
- FRISCHHOLZ, R. W., DIECKMANN, U. BioID: A Multimodal Biometric Identification System, *IEEE Computer*, v. 33, n. 2, February 2000, p. 64-68.
- GATECH. Disponível em: <ftp://ftp.cc.gatech.edu/pub/gvu/cpl/walkers/subjects>. Acesso em: 20 mar.2006.
- GONZALEZ, R. C.; WOODS, R. E. Digital Image Processing, *Prentice Hall*, 2nd Edition, 2002.
- HAN, J.; BHANU, B. Individual Recognition Using Gait Energy Image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 28, n. 2, February, 2006, p. 316-322.
- HAN, J.; BHANU, B. Human Activity Recognition in thermal infrared imagery. *IEEE Computer Society Conference on Vision and Pattern Recognition*, v. 3, 2005, p. 17-25.
- HANSEN, L. K.; LARSEN, J.; KOLENDA, T. Blind Detection of Independent Dynamic Components. *In proc. IEEE ICASSP'2001*, v. 1, 2001, p. 3197-3200.
- HAYFRON-ACQUAH, J. B.; NIXON, M. S.; CARTER, J. N. Automatic Gait Recognition by Symmetry Analysis. *Pattern Recognition Letters* 24, Elsevier Science B.V, 2003, p. 2175-2183.
- HAYKIN, S. Neural Networks a Comprehensive Foundation. *Macmillan College Publishing Company*, New York, 1994, p. 236-281.
- HYVÄRINEN, A.; KARHUNEN, J.; OJA, E. *Independent Component Analysis*. New York: John Wiley & Sons, 2001.
- HONG, L.; JAIN, A. Integrating Faces and Fingerprints for Personal Identification. *IEEE Trans. Pattern Analysis and Machine Intelligence*, v. 20, n. 12, 1998, p. 1295-1307.
- HUANG, P.S.; Harris, C.J.; Nixon, M.S. Recognizing Humans by Gait using a Statistical Approach for Temporal Templates. *Artificial Intelligence in Engineering*, 13(4), October 1998, p.4556-4561.
- INDOVINA, M.; ULUDAG, U.; SNELICK, R.; MINK, A.; JAIN, A. Multimodal Biometric Authentication Methods: a COTS Approach. *Proc. of Workshop on Multimodal User Authentication*, Santa Barbara, CA, 2003, p. 99-106.
- JAIN, A. K.; ROSS, A. Multibiometric Systems. *Communications of the ACM*, v. 47, n 1, 2004.
- JAIN, A. K.; ROSS, A.; PRABHAKAR, S. An Introduction to Biometric Recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, v. 14, n. 1, 2004, p. 4-19.

- KALE, A.; RAJANAGOPALAN, A. N.; CUNTOOR N.; KRÜGER, V. Gait-based recognition of humans using continuous HMMs. *Proceedings of the fifth IEEE International Conference on Automatic Face and Gestures Recognition*, 2002.
- KALE, A.; ROY-CHOWDHURY, A.; CHELLAPPA, R. Fusion of Gait and Face for Human Identification. *Int. Conf. on Acoustics, Speech and Signal Processing*, v. 5, 2002, p. 901-904.
- KOZLOWSKI, L.; CUTTING, J. Recognizing the Sex of a Walker from a Dynamic Point Display. *Perception and Psychophysics*, n. 21, 1977, p. 575–580.
- LARSEN, J.; ANDERSEN, L. N.; HINTZ-MADSEN, M.; HANSEN, L.K., Design of Robust Neural Network Classifiers. *Proceedings of the 1998 International Conference on Acoustics, Speech and Signal Processing*, New York. 1998, p. 1205-1208.
- LEE, K-C.; HO, J.; YANG M-H.; KRIEGMAN, D. Video-based Face Recognition using Probabilistic Appearance Manifolds. In *Proc. of IEEE Int. Conf. on Computer Vision and Pattern Recognition*, v. 1, Illinois, USA, 2003, p. 313–320.
- LEE, L.; GRIMSON, W.E.L. Gait Analysis for Recognition and Classification. *Proceedings of the IEEE Conference on Face and Gesture Recognition*, 2002, p. 155–161.
- LITTLE, J. J.; BOYD, J. E. Recognizing People by Their Gait: The Shape of Motion. *Videre: Journal of Computer Vision Research*, Quarterly Journal, v.1, n. 2, The MIT Press, Cambridge, Massachusetts, 1998.
- MELATO, L.; BIGAL, M. E.; SPECIAL, J. G. Hidrocefalia de Pressão Normal: Avaliação de Cinco Anos de Experiência e Revisão de Literatura. *Medicina*, Ribeirão Preto, n. 33, 2000, p. 499-505.
- NIELSEN, H.B. UCMINF - An Algorithm for Unconstrained, Nonlinear Optimization. *IMM-TEC-0019*, IMM, Technical University of Denmark, 2001.
- NIXON, M. S.; CARTER, J. N.; CUNADO, D.; HUANG, P. S.; STEVENAGE., S. V. Automatic Gait Recognition. In: *Anil K. Jain, Ruud Bolle and Sharath Pankanti Eds.: Biometrics: Personal Identification in Networked Society*, 1999, p. 231-250.
- NOVACHEK, T. F. The Biomechanics of Running. *Gait and Posture 7 - Elsevier Science B.V.*, 1998, p. 77–95.
- OLIVETTI RESEARCH LTD: Face Database. Disponível em: <<http://www.cam-orl.co.uk/facedatabase.html>>. Acesso em: 06 set.2005.
- PHILLIPS, P. J.; SARKAR, S.; ROBLEDO, I.; GROTH, P; BOWYER, K. Baseline Results for the Challenge Problem of Human ID Using Gait Analysis. *Proc. of the 5th IEEE International Conference on Automatic Face and Gesture Recognition*, FG. '02, Washington, D.C, 2002, p.137 – 143.
- ROSS, A.; JAIN, A. K. Information Fusion in Biometrics. *Pattern Recognition Letters*, n. 24, 2003, p. 2115-2125.

- ROSS, A.; JAIN, A. K. Multimodal Biometrics: an Overview, *Proc. of 12th European Signal Processing Conference (EUSIPCO)*, Austria, 2004, p.1221-1224.
- SALOMÃO, J. M.; ALMEIDA, A. R.; SALLES, E. O. T. Avaliação de Desempenho de um Classificador SVM no Reconhecimento Pessoal via Forma de Caminhar com Seleção de Características através das Transformadas Wavelets. *VII Simpósio Brasileiro de Automação Inteligente*, São Luis-MA, 1 CD-ROM, set. 2005a.
- SALOMÃO, J. M.; ALMEIDA, A. R.; SALLES, E. O. T. Avaliação das Características da Forma de Caminhar na Identificação Pessoal com o Apoio da Análise de Componentes Independentes. *Proceedings of the XXVI Iberian Latin-American Congress on Computational Methods in Engineering*, Guarapari-ES, 1 CD-ROM, 19-21 out. 2005b.
- SALOMÃO, J. M.; ALMEIDA, A. R.; SALLES, E. O. T. Reconhecimento Pessoal através da Fusão Hierárquica das Características da Forma de Caminhar e de Imagens Faciais. *XVI Congresso Brasileiro de Automática*, Salvador-BA, 1 CD-ROM, 03-06 out. 2006a.
- SALOMÃO, J. M.; ALMEIDA, A. R.; SALLES, E. O. T. Uso das Características Biomecânicas do Caminhar na Identificação Pessoal através de Redes RBF. *XX Congresso Brasileiro de Engenharia Biomédica*, São Pedro-SP, 1 CD-ROM, 22-26 out. 2006b.
- SCHOLHORN, W. I.; NIGG B.M; STEPHANSHYN, D.J; LIU, W. Identification of Individual Walking Patterns using Time Discrete and Time Continuous Data Sets. *Gait and Posture*, v. 15, 2002, p. 180–186.
- SHAKHNAROVICH, G.; LEE, L.; DARRELL, T. Fusion of Face and Gait Evidence for Recognition with Multiple Views. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition*, Lihue, HI, 2001, p. 439-446.
- SHAKHNAROVICH, G.; DARRELL, T. On Probabilistic Combination o Face and Gait Cues for Identification. *Proc. 5th IEEE International Conference on Face and Gesture Recognition*, USA, 2002, p. 169-174.
- SIGURDSSON, S.; LARSEN, J.; HANSEN, L.K; PHILPSEN, P.A.; WULF, H.C., Outlier Estimation and Detection: Application to Skin Lesion Classification. *The 27th International Conference on Acoustic, Speech, and Signal Processing*, v.1, 2002, p. 1049-1052.
- SINGH, S.; GYAOUROVA, A.; BEBIS, G.; PAVLIDIS, I., Infrared and Visible Image Fusion for Face Recognition. *Proceedings of the SPIE Defense and Security Symposium*, Orlando, Florida, v. 5404, April 12-16, 2004, p. 585-596.
- SOCOLINSKY, D. A; SELINGER, A. Thermal Face Recognition in an Operational Scenario, *Conference on Computer Vision and Pattern Recognition n..2*, 2004, p. 1012-1019.
- SOCOLINSKY, D.; SELINGER, A. A Comparative Analysis of Face Recognition Performance with Visible and Thermal Infrared Imagery. *ICPR (4) Proceedings of the International Conference on Pattern Recognition*, 2002.

- SOCOLINSKY, D.; WOLFF, L.; NEUHEISEL, J. D.; EVELAND, C. K., Illumination Invariant Face Recognition Using Thermal Infrared Imagery. *CVPR (1) Proceedings of the International Conference on Pattern Recognition*, 2001, p. 527-534
- SOCOLINSKY, D.; NEUHEISEL, J. D.; PRIEBE, C. E.; DEVINNEY, J.; MARCHETTE, D. Fast Face Detection with a Boosted CCD Classifier. *Computing Science and Statistics*, 2003, n.. 35.
- STEVENAGE., S. V.; NIXON, M. S.; VINCE., K. Visual Analysis of Gait as a Cue to identity. *Applied Cognitive Psychology*, 13, New York, 1999, p. 513-526.
- TAN, T; YAN, H. Face Recognition using the Weighted Fractal Neighbor Distance. *IEEE Transactions on Systems, Man and Cybernetics*, Part C, v. 35, no. 4, 2005, p. 576-582.
- TOLLIVER, D.; COLLINS, R. Gait shape estimation for identification, *Proceedings of AVBPA*, p. 734-742, 2003.
- TURK, M. A.; PENTLAND, A. P., Eigenfaces for Recognition, *J. Cognitive Neurosci.* 3 (1) 71-86. 1991.
- YAM, C.-Y.; NIXON, M. S.; CARTER, J.N. Gait Recognition by Walking and Running: A Model-based Approach. *In: Proc. 5th Asian Conf. Comput. Vision*, 2002, p.1-6.
- WANG, L.; NING, H.; TAN, T.; HU, W. Fusion of Static and Dynamic Body Biometrics for Gait Recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 2004, v. 14, p. 149-158.
- WANG, Y.; TAN,T.; JAIN, K.A. Combining Face and Iris Biometrics for Identity Verification. *Proc. of 4th International Conference on AVBPA*, 2003.
- WANG, L.; TAN, T.; NING, H.; HU, W. Silhouette Analysis-Based Gait Recognition for Human Identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, v. 25, n. 12.
- ZHAO, W.; CHELLAPPA, R.; PHILLIPS, P.; ROSENFELD, A. Face Recognition: A Literature Survey. *ACM Computing Surveys*, December 2003, v. 35, n. 4, p. 399–458.